Simulation based forward modeling of galaxy populations

Alex Alarcon
Junior Leader Fellow 'la Caixa'
Institute of Space Sciences (ICE-CSIC)

"la Caixa" Foundation

Institute of Space Sciences

CSIC Consejo Superior de Investigaciones Científicas IEEC

Simulation based forward modeling of galaxy populations

Alex Alarcon
Junior Leader Fellow 'la Caixa'
Institute of Space Sciences (ICE-CSIC)

"la Caixa" Foundation

Institute of Space Sciences
CSIC IEEC

In collaboration with:

Andrew Hearin

Gillian Beltz-Mohrmann

Matt Becker

Jonas Chaves-Montero

Alan Pearl

Differentiable sky predictions

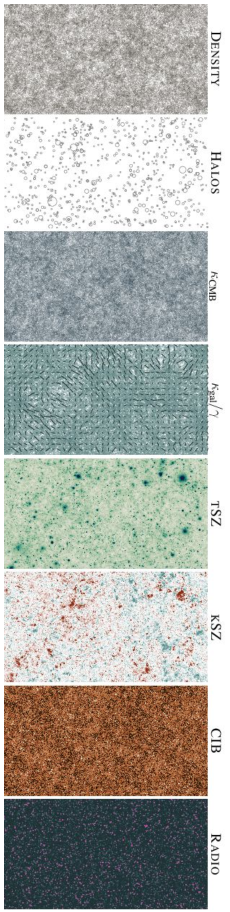
Diffmah (halo mass)
Diffprof (halo concentration)

Diffstar (galaxy in-situ star formation history)
Diffburst (galaxy starbursts)

DSPS (composite stellar populations)
Diffdust (dust attenuation)

Diffmerge (galaxy merging)

Multi-λ predictions



Inset image credit: Millennium XXL simulation, NASA, ESA, Yuuki Omori/Agora simulation

Differentiable sky predictions

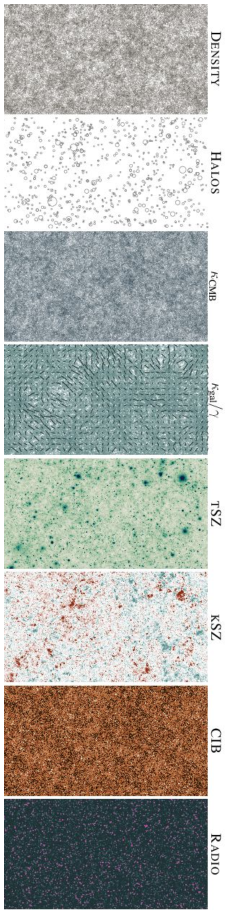
Diffmah (halo mass)
Diffprof (halo concentration)

Diffstar (galaxy in-situ star formation history)
Diffburst (galaxy starbursts)

DSPS (composite stellar populations)
Diffdust (dust attenuation)

Diffmerge (galaxy merging)

Multi-λ predictions



- Building **ground-up reformulation** of galaxy–halo connection.

Inset image credit: Millennium XXL simulation, NASA, ESA, Yuuki Omori/Agora simulation

Differentiable sky predictions

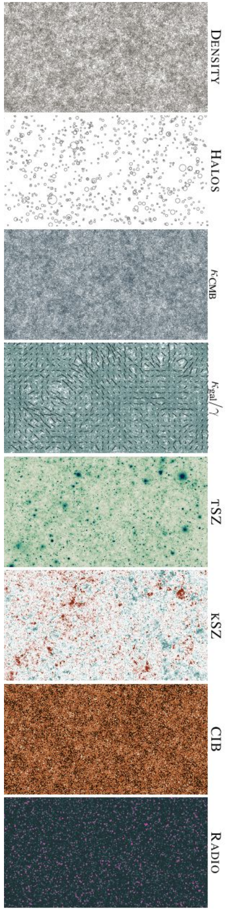
Diffmah (halo mass)
Diffprof (halo concentration)

Diffstar (galaxy in-situ star formation history)
Diffburst (galaxy starbursts)

DSPS (composite stellar populations)
Diffdust (dust attenuation)

Diffmerge (galaxy merging)

Multi-λ predictions



- Building **ground-up reformulation** of galaxy–halo connection.
- We seek to identify the **minimum but interpretable** parametric flexibility that is required to accurately capture the data.

Inset image credit: Millennium XXL simulation, NASA, ESA, Yuuki Omori/Agora simulation

Differentiable sky predictions

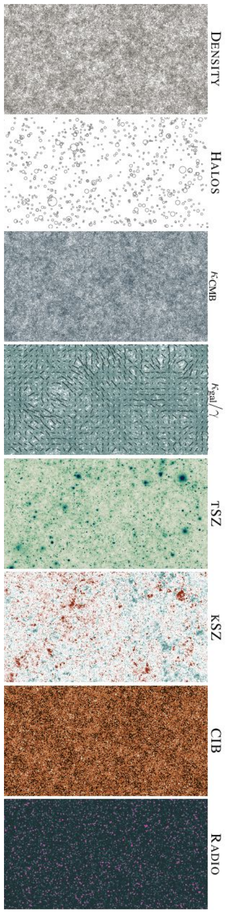
Diffmah (halo mass)
Diffprof (halo concentration)

Diffstar (galaxy in-situ star formation history)
Diffburst (galaxy starbursts)

DSPS (composite stellar populations)
Diffdust (dust attenuation)

Diffmerge (galaxy merging)

Multi-λ predictions



- Building **ground-up reformulation** of galaxy–halo connection.
- We seek to identify the **minimum but interpretable** parametric flexibility that is required to accurately capture the data.
- Models become **automatically differentiable** thanks to their implementation in JAX.

Inset image credit: Millennium XXL simulation, NASA, ESA, Yuuki Omori/Agora simulation

Differentiable sky predictions

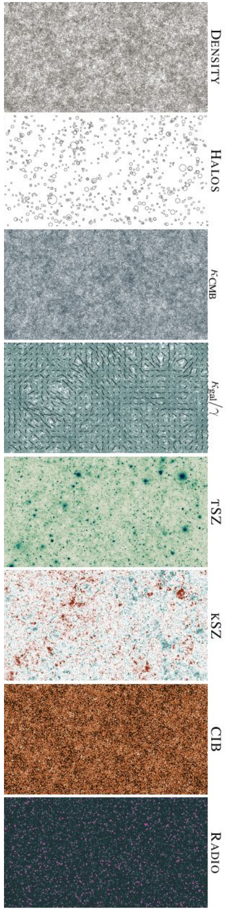
Diffmah (halo mass)
Diffprof (halo concentration)

Diffstar (galaxy in-situ star formation history)
Diffburst (galaxy starbursts)

DSPS (composite stellar populations)
Diffdust (dust attenuation)

Diffmerge (galaxy merging)

Multi-λ predictions



Inset image credit: Millennium XXL simulation, NASA, ESA, Yuuki Omori/Agora simulation

Diffmah: 2105.05859

Diffmah: Halo Mass Assembly

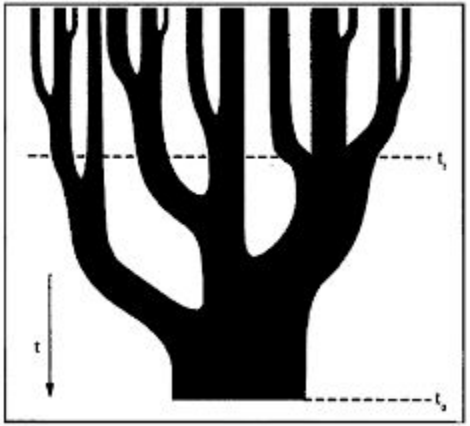
Diffmah is a parametric model for halo mass assembly.

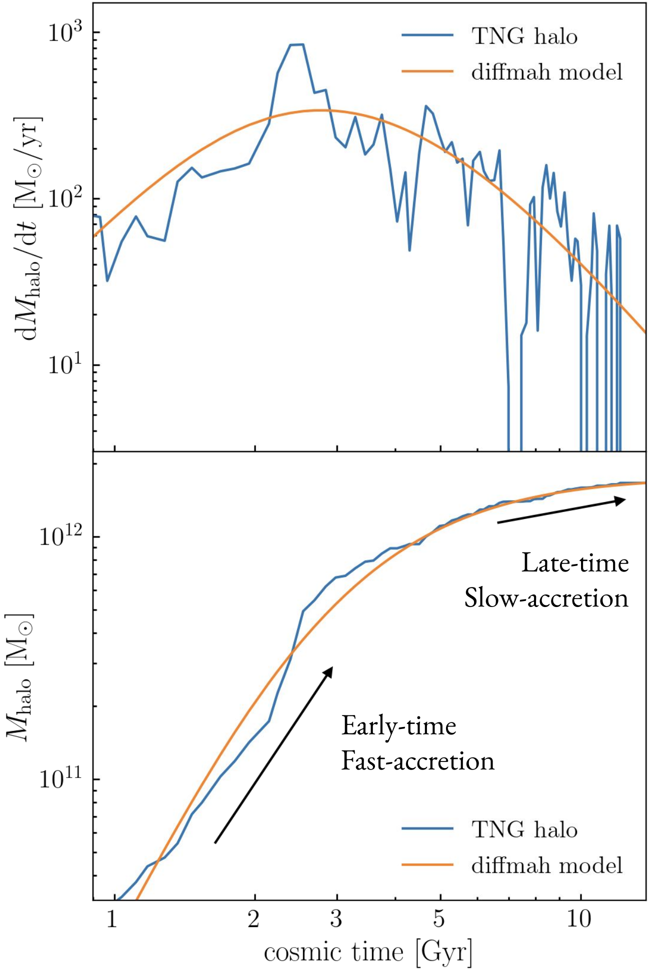
The base is a high resolution N-body simulation with merger trees.

Each halo in the simulation is then replaced by a best-fitting 4 parameter formula:

$$M_{\rm halo}(t) = M_0(t/t_0)^{\alpha(t)}$$

Power law with a rolling index





Differentiable sky predictions

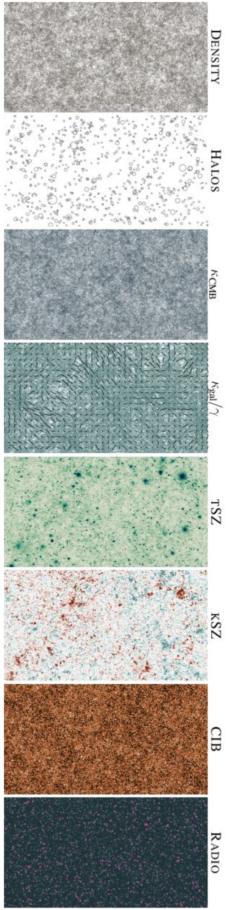
Multi-λ predictions

Diffmah (halo mass)
Diffprof (halo concentration)

Diffstar (galaxy in-situ star formation history)
Diffburst (galaxy starbursts)

DSPS (composite stellar populations)
Diffdust (dust attenuation)

Diffmerge (galaxy merging)



Inset image credit: Millennium XXL simulation, NASA, ESA, Yuuki Omori/Agora simulation

Diffstar: Gas and halo mass assembly

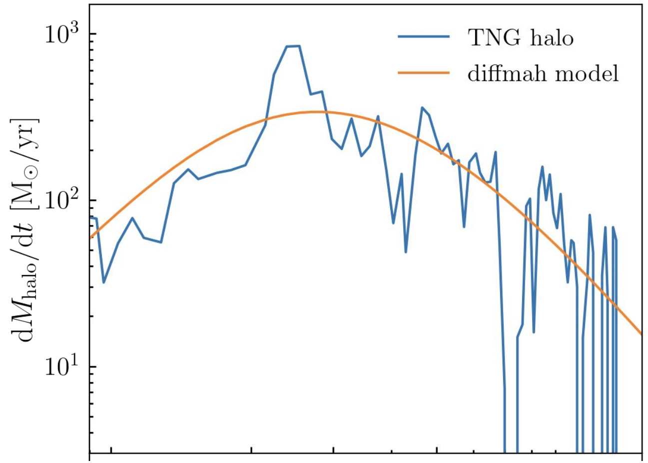
The Diffstar model assumes that baryonic matter becomes available for star formation at a rate that is closely related to the growth rate of the dark matter halo:

$$\frac{\mathrm{d}M_\mathrm{g}}{\mathrm{d}t} = f_\mathrm{b}\frac{\mathrm{d}M_\mathrm{h}}{\mathrm{d}t}$$

Gas accretion rate → $\frac{\mathrm{d}M_\mathrm{g}}{\mathrm{d}t}$

cosmic baryon fraction → f_b

Mass accretion rate → $\frac{\mathrm{d}M_\mathrm{h}}{\mathrm{d}t}$



Diffstar: 2205.04273

Diffstar: Gas and halo mass assembly

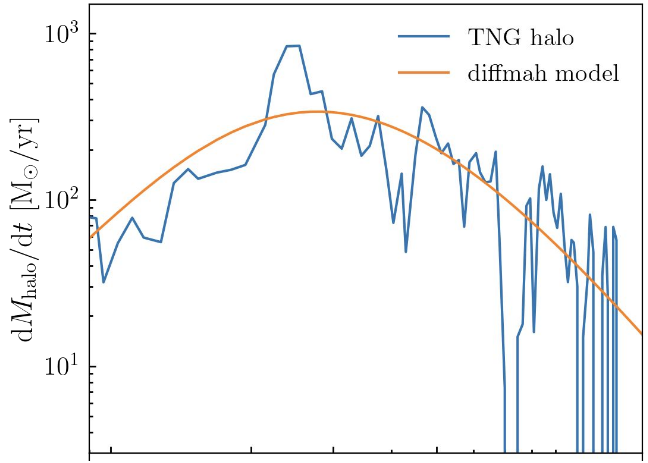
The Diffstar model assumes that baryonic matter becomes available for star formation at a rate that is closely related to the growth rate of the dark matter halo:

$$\frac{\mathrm{d}M_{\mathrm{g}}}{\mathrm{d}t} = f_{\mathrm{b}} \frac{\mathrm{d}M_{\mathrm{h}}}{\mathrm{d}t}$$

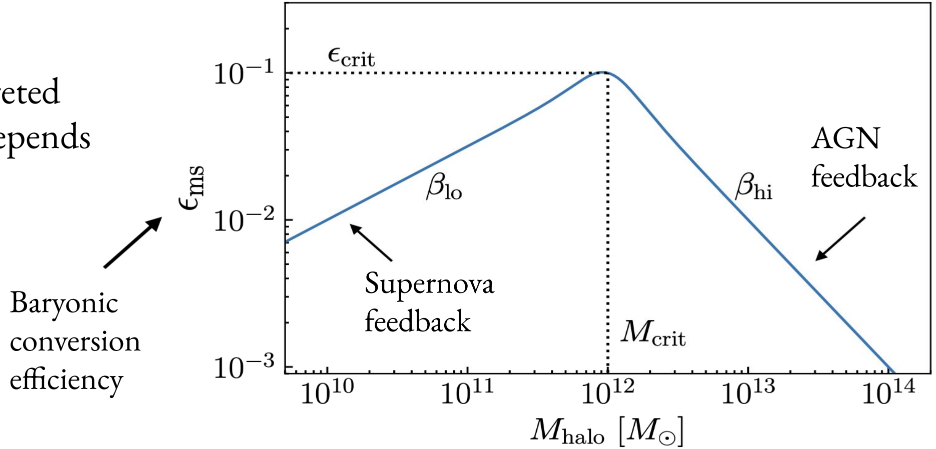
Gas accretion rate

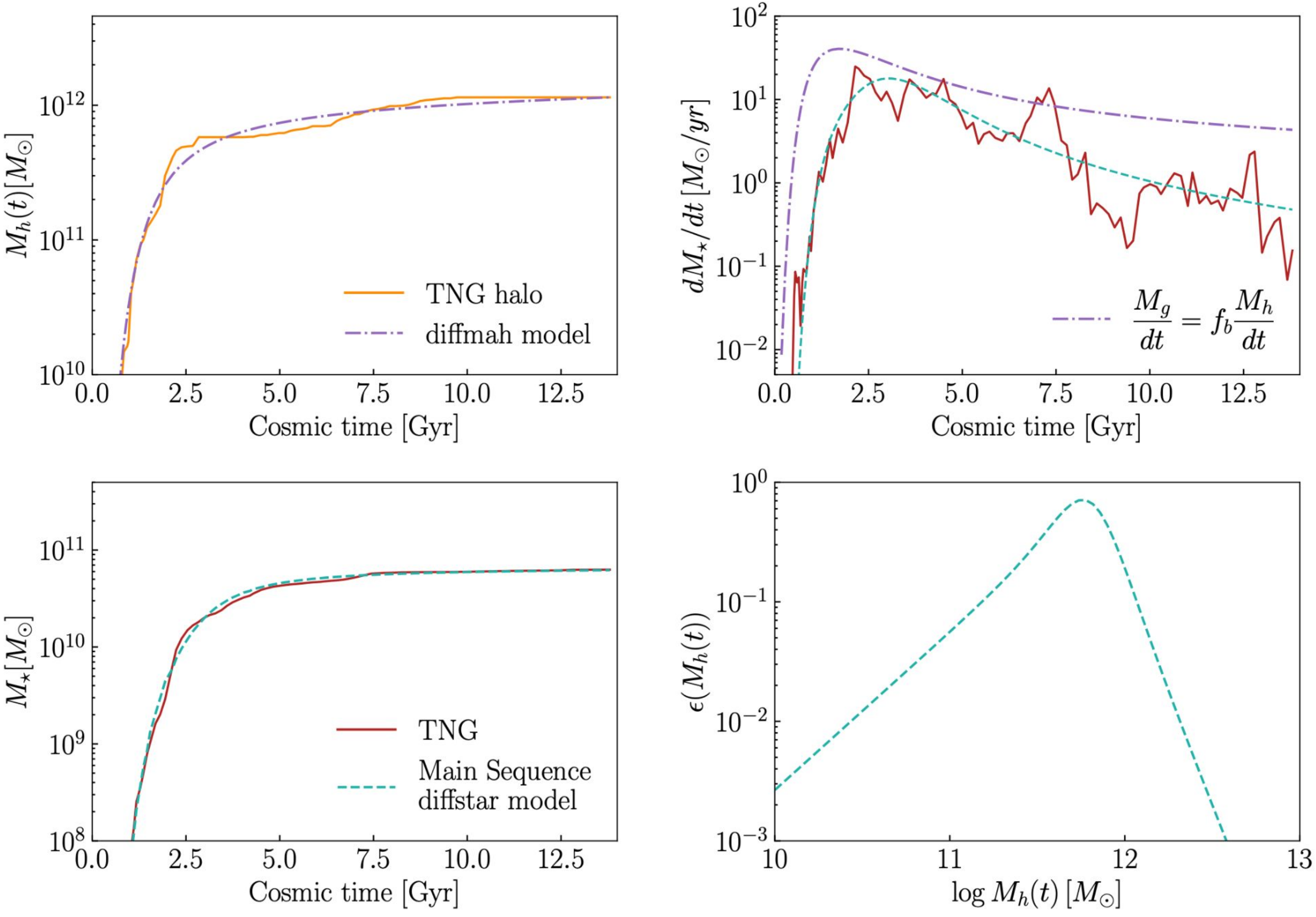
Mass accretion rate

cosmic baryon fraction

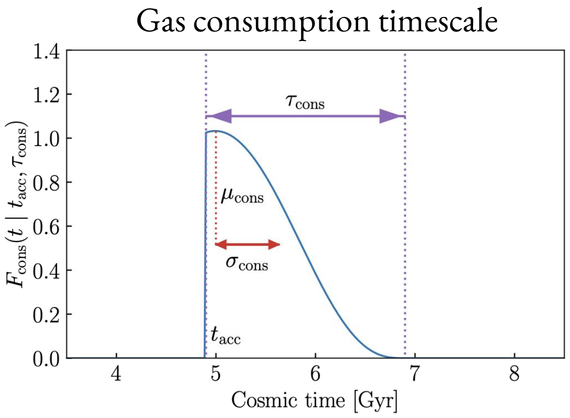


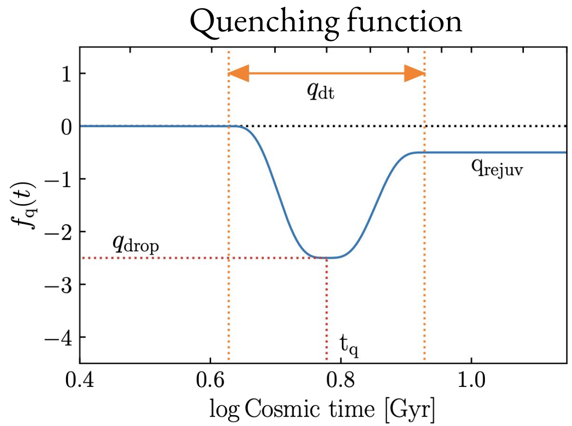
The Diffstar model assumes that only a fraction of the accreted material ever transforms into stars, and that this fraction depends only upon the instantaneous mass of the parent halo.

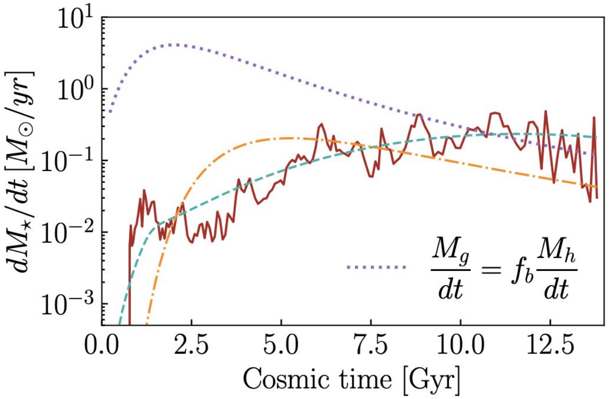


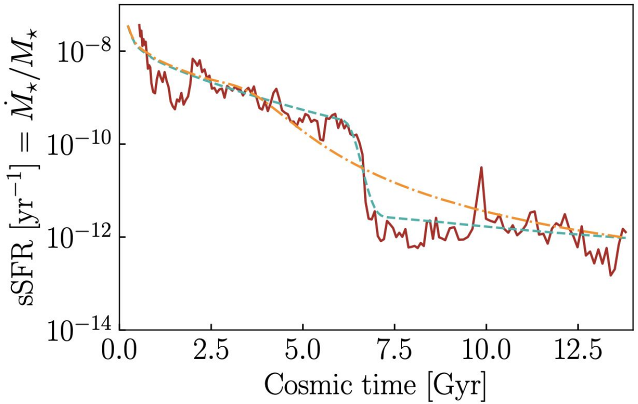

$M_h(t)[M_\odot]$
10^{10}
10^{11}
10^{12}
0.0
2.5
5.0
7.5
10.0
12.5
Cosmic time [Gyr]
TNG halo
diffmah model
$dM_\star/dt\,[M_\odot/yr]$
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
0.0
2.5
5.0
7.5
10.0
12.5
Cosmic time [Gyr]
$\frac{M_g}{dt} = f_b \frac{M_h}{dt}$
$M_\star[M_\odot]$
10^{8}
10^{9}
10^{10}
10^{11}
0.0
2.5
5.0
7.5
10.0
12.5
Cosmic time [Gyr]
TNG
Main Sequence
diffstar model
$\epsilon(M_h(t))$
10^{-3}
10^{-2}
10^{-1}
10^{0}
10
11
12
13
$\log M_h(t)\,[M_\odot]$

Diffstar: Gas consumption and Quenching





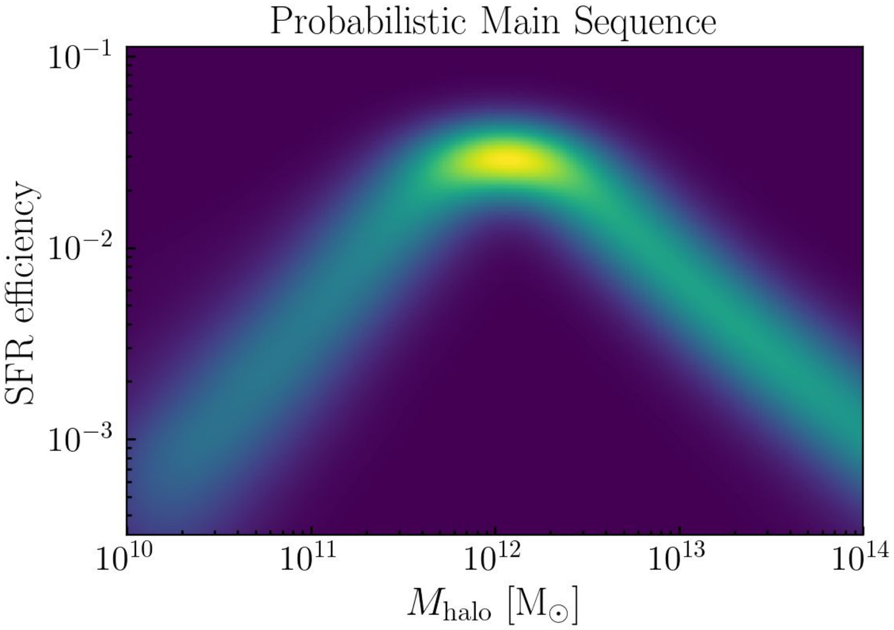




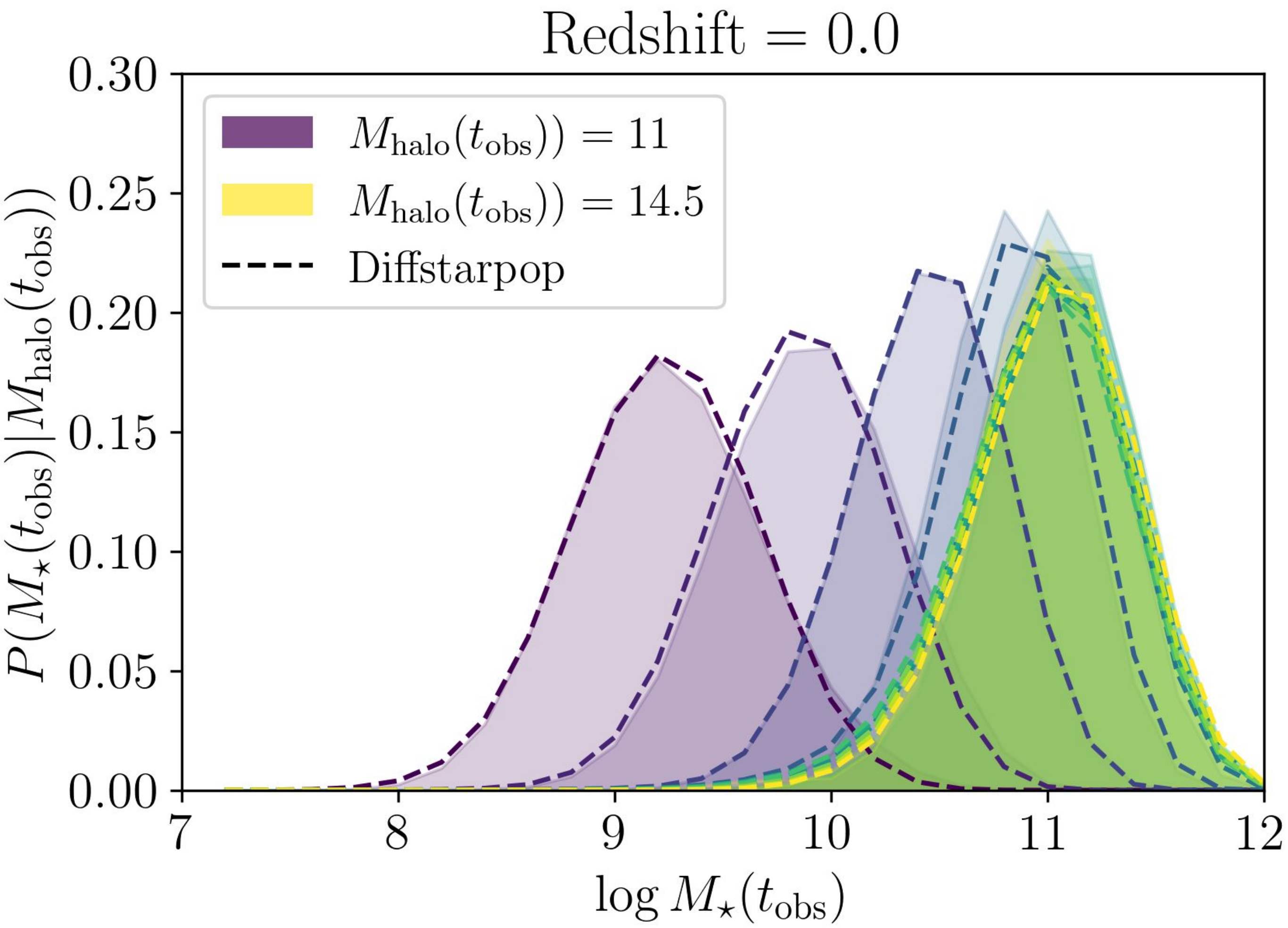
DiffstarPop: Properties of galaxy populations

$$\theta_{\mathrm{SFH}} \rightarrow \mathrm{SFH(t)}$$

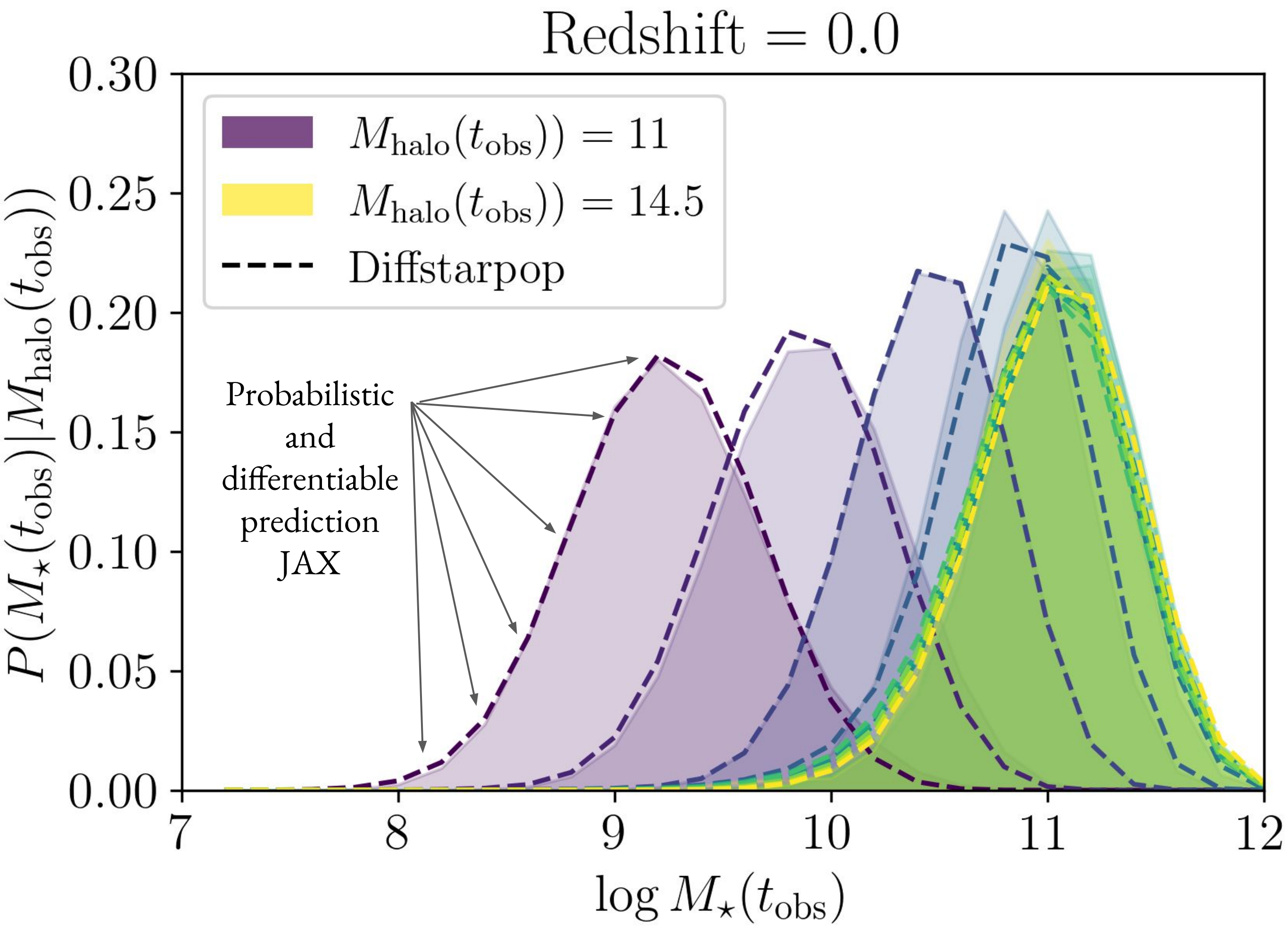
$$\Psi_{\mathrm{SFH}} \rightarrow P(\theta_{\mathrm{SFH}} | \Psi_{\mathrm{SFH}}) \rightarrow P(\mathrm{SFH(t)})$$



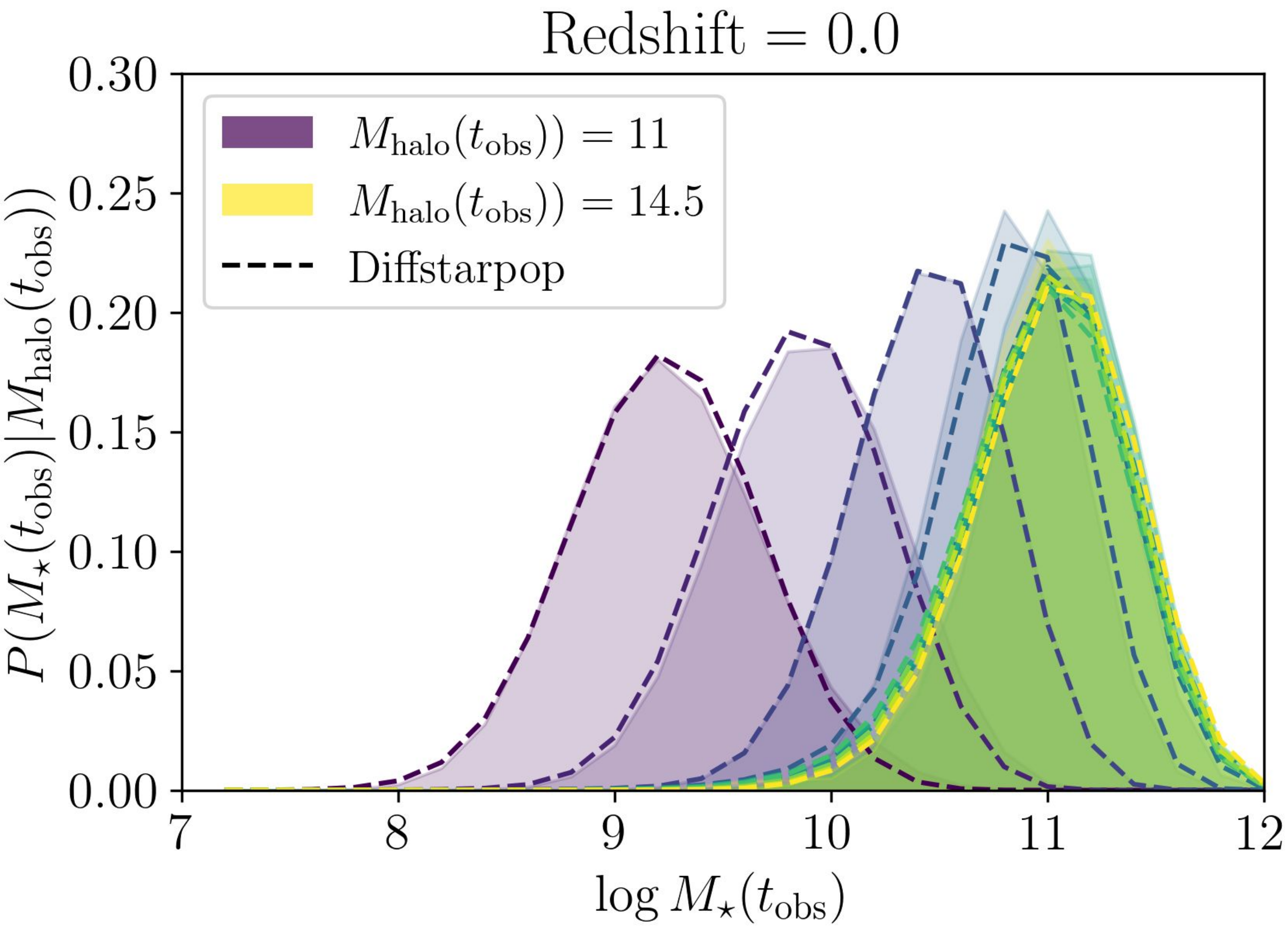
Parametrize a *probabilistic* galaxy that lives in each halo



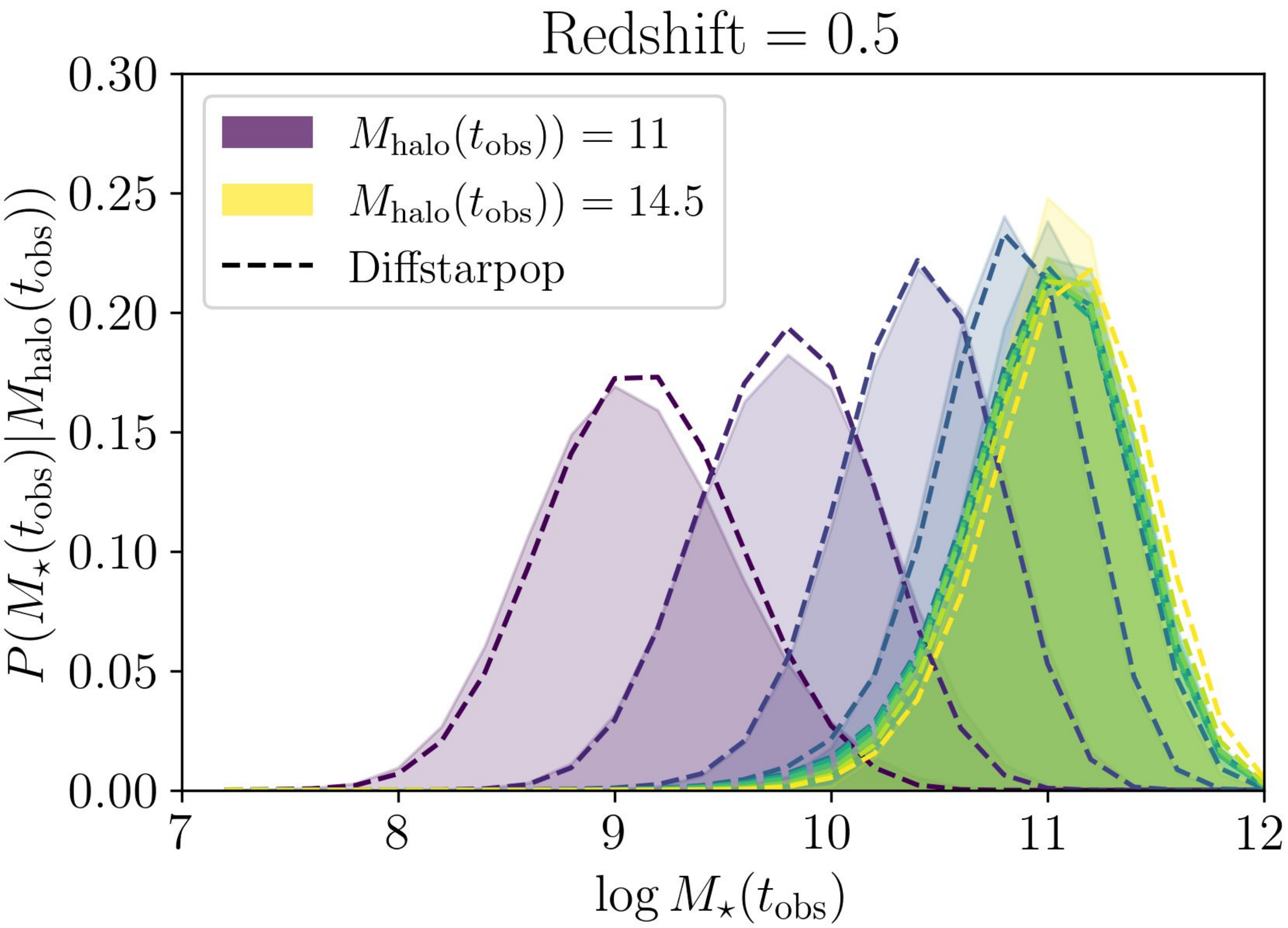
Redshift = 0.0
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
$\log M_\star(t_{\rm obs})$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
7
8
9
10
11
12



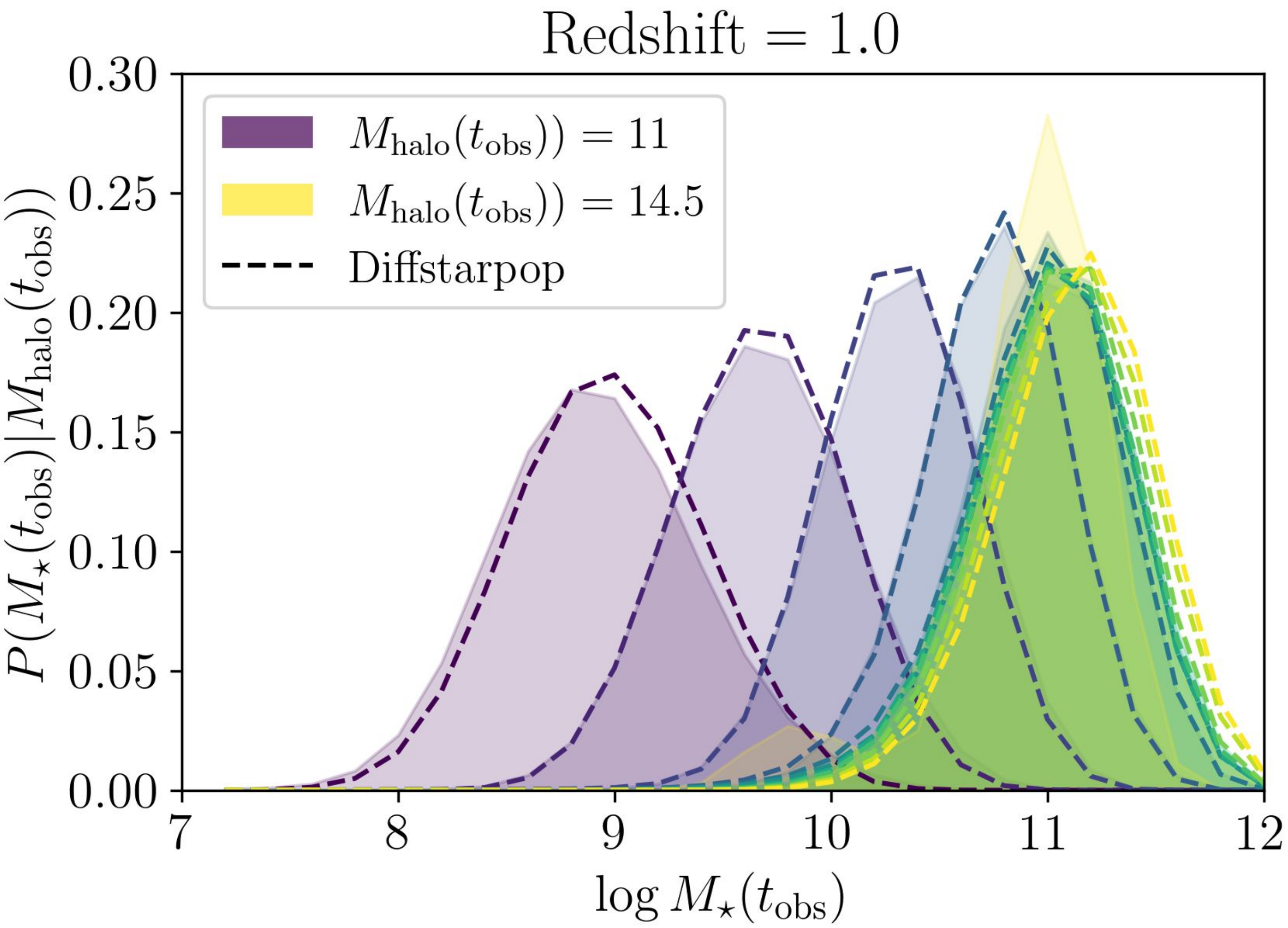
Redshift = 0.0
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
Probabilistic and differentiable prediction JAX
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
$\log M_\star(t_{\rm obs})$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
7
8
9
10
11
12



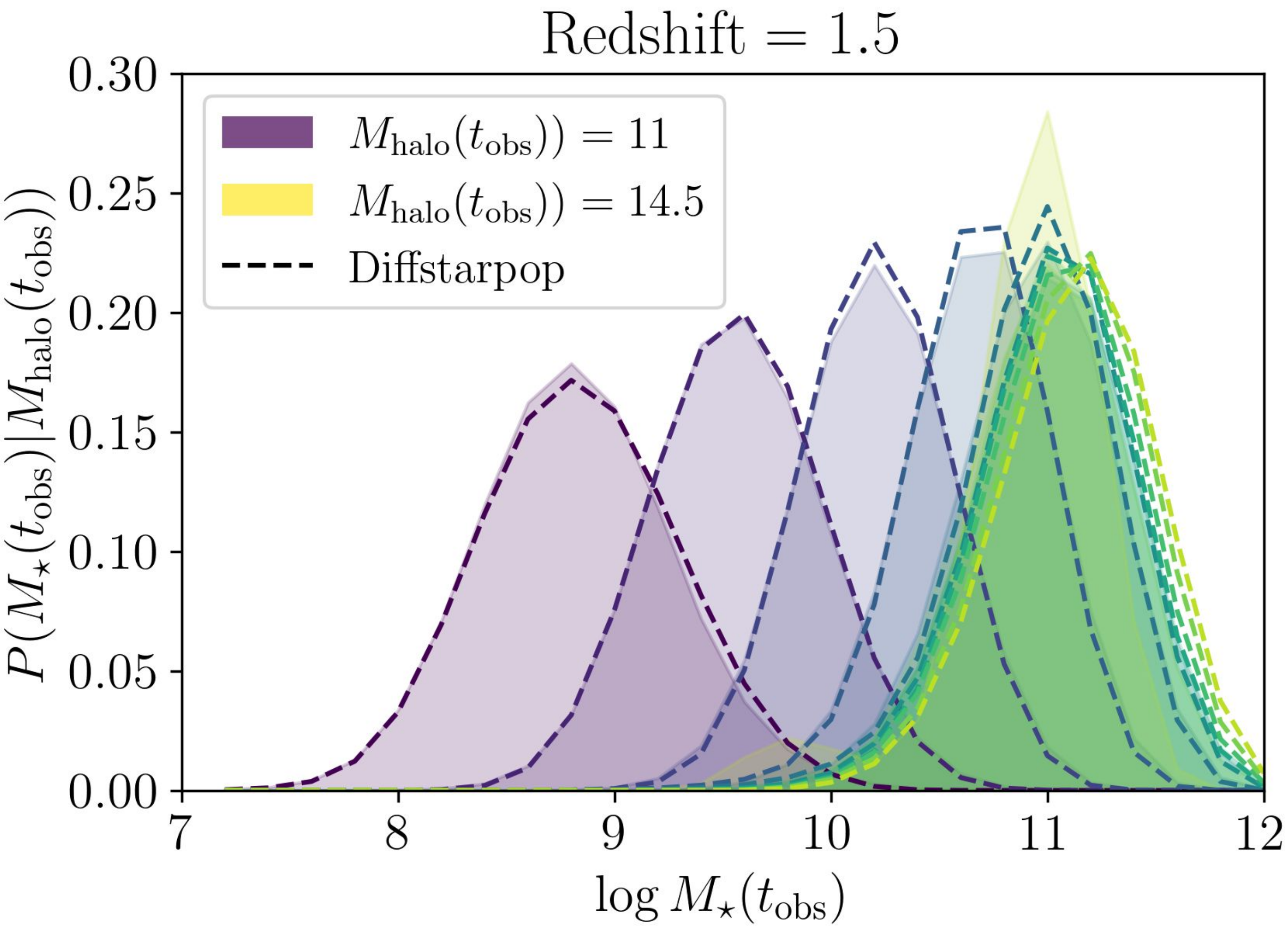
Redshift = 0.0
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
7
8
9
10
11
12
$\log M_\star(t_{\rm obs})$

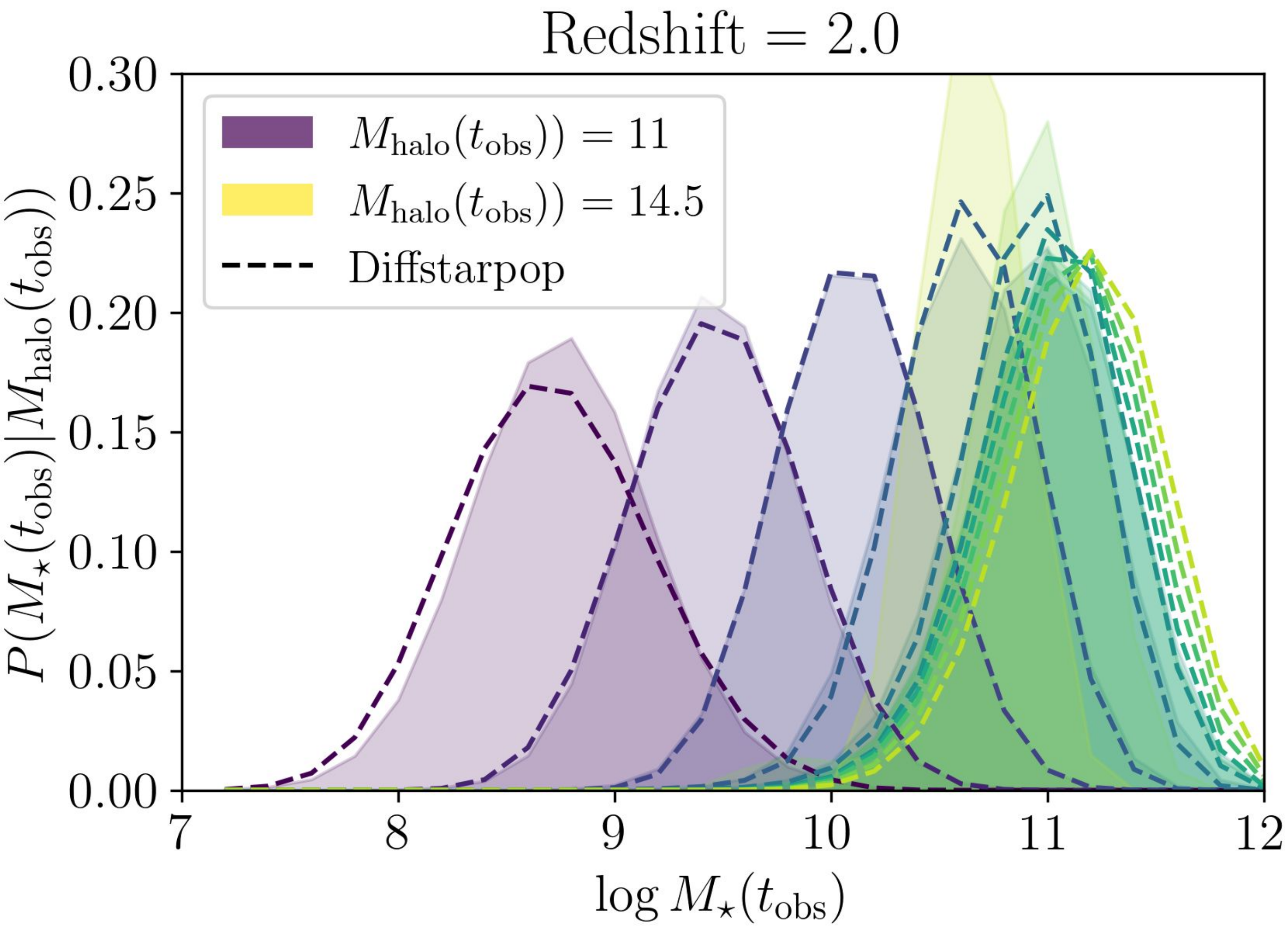


Redshift = 0.5
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
7
8
9
10
11
12
$\log M_\star(t_{\rm obs})$

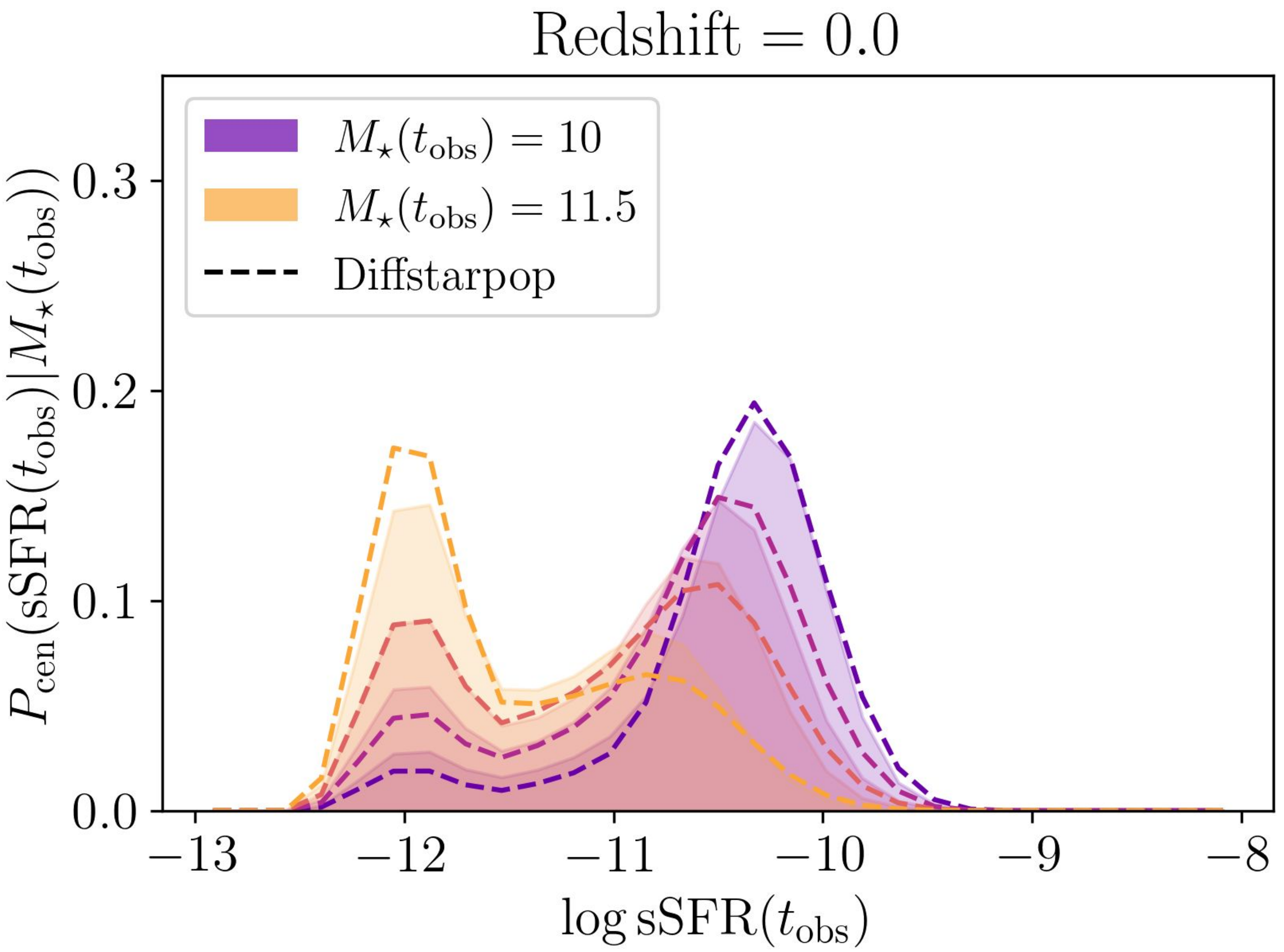


Redshift = 1.0
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
$\log M_\star(t_{\rm obs})$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
7
8
9
10
11
12

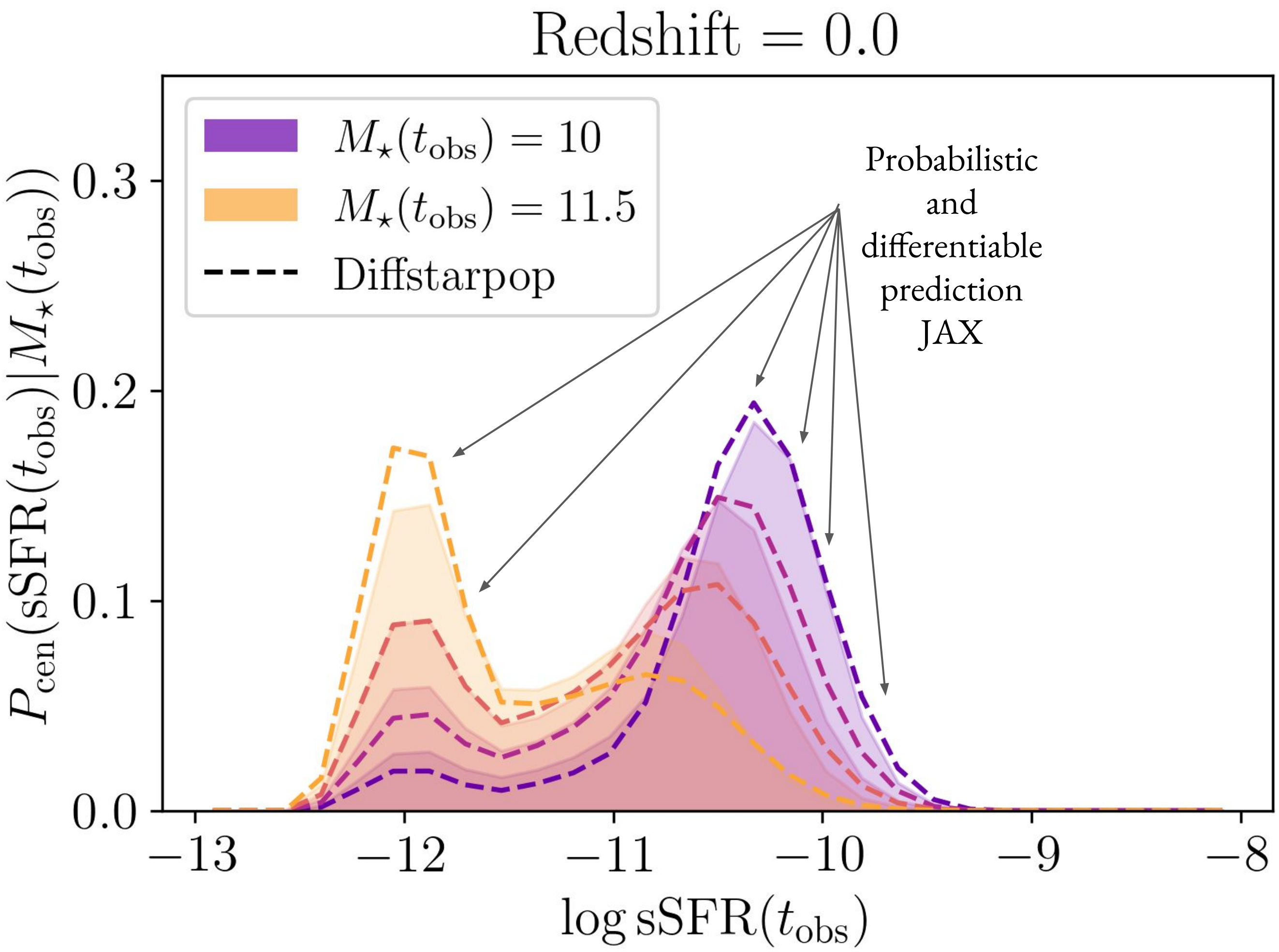

Redshift = 1.5
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
$\log M_\star(t_{\rm obs})$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
7
8
9
10
11
12



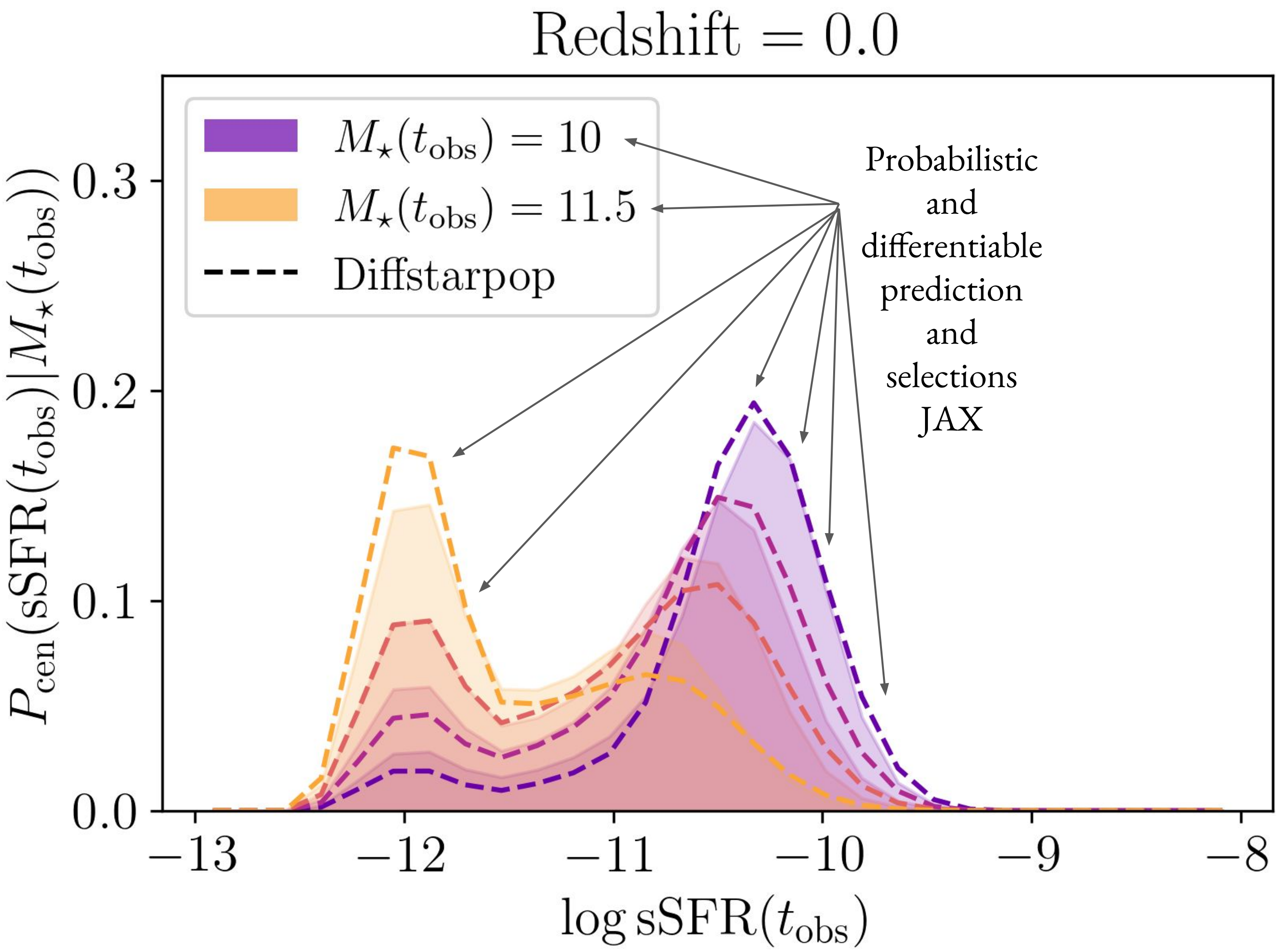
Redshift = 2.0
$M_{\rm halo}(t_{\rm obs})) = 11$
$M_{\rm halo}(t_{\rm obs})) = 14.5$
Diffstarpop
$P(M_\star(t_{\rm obs})|M_{\rm halo}(t_{\rm obs}))$
$\log M_\star(t_{\rm obs})$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
7
8
9
10
11
12



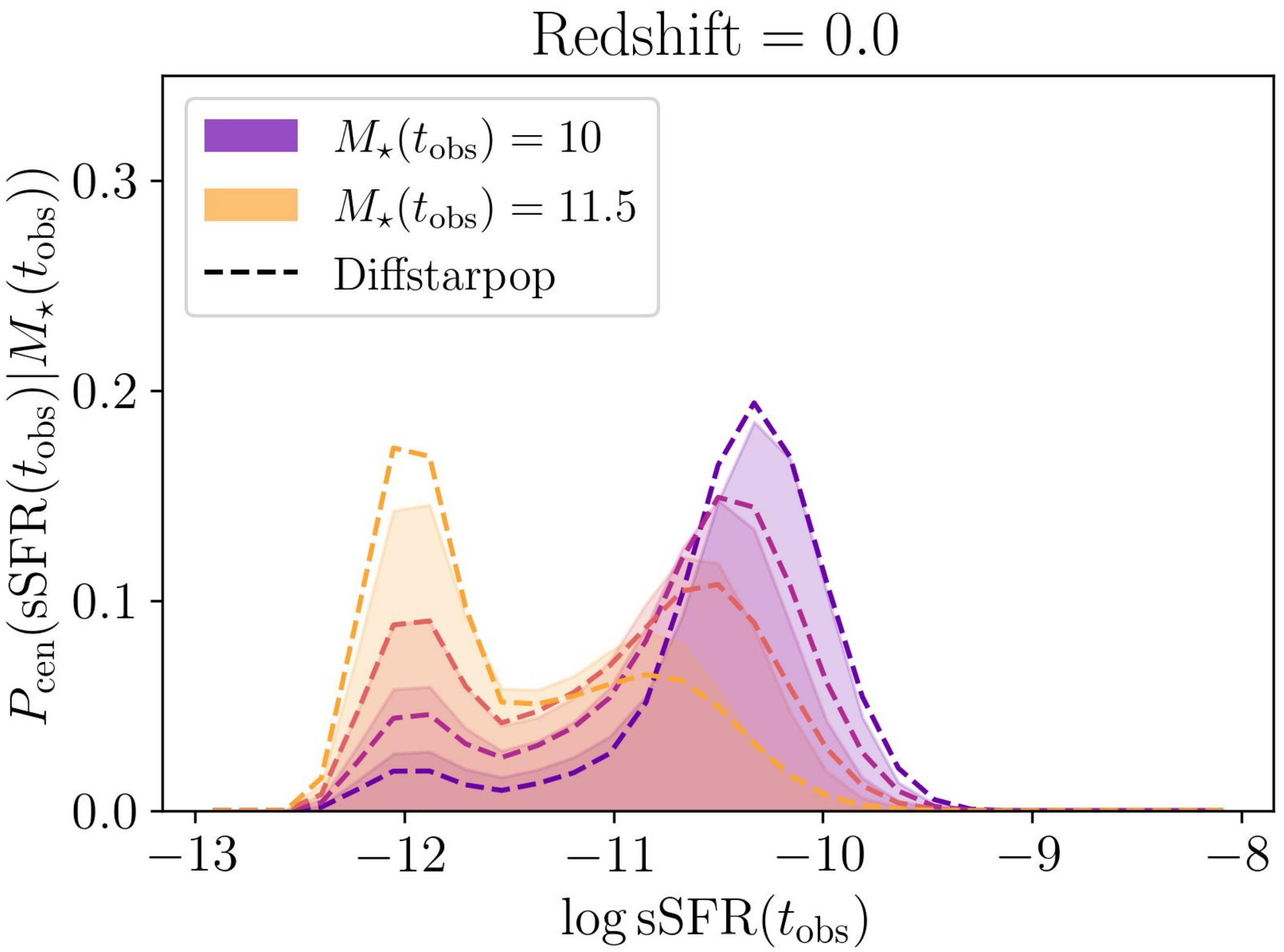
Redshift = 0.0
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
$P_{\rm cen}(\mathrm{sSFR}(t_{\rm obs})|M_\star(t_{\rm obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \mathrm{sSFR}(t_{\rm obs})$



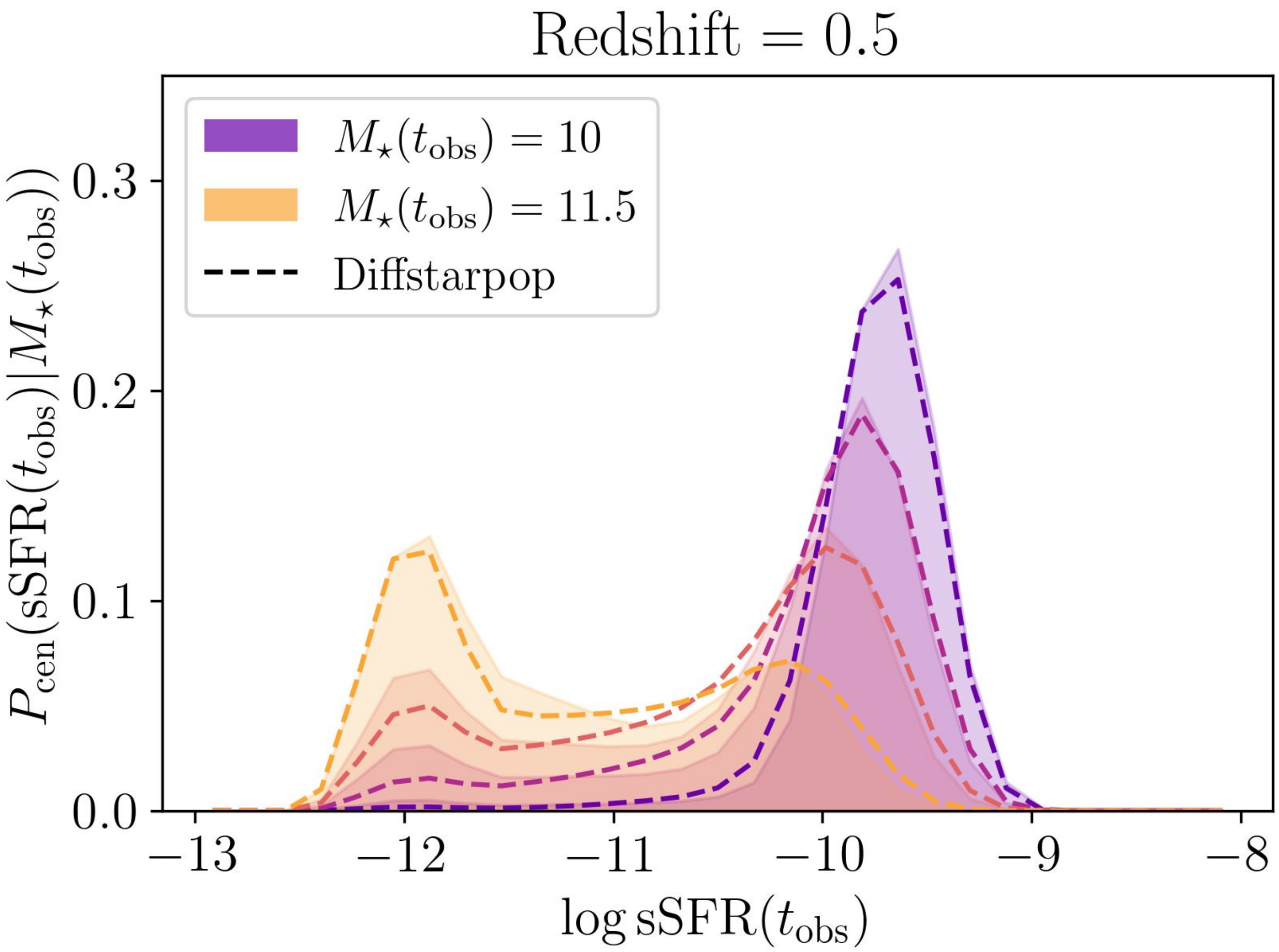
Redshift = 0.0
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
Probabilistic
and
differentiable
prediction
JAX
$P_{\rm cen}(\rm sSFR(t_{obs})|M_\star(t_{obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \rm sSFR(t_{obs})$



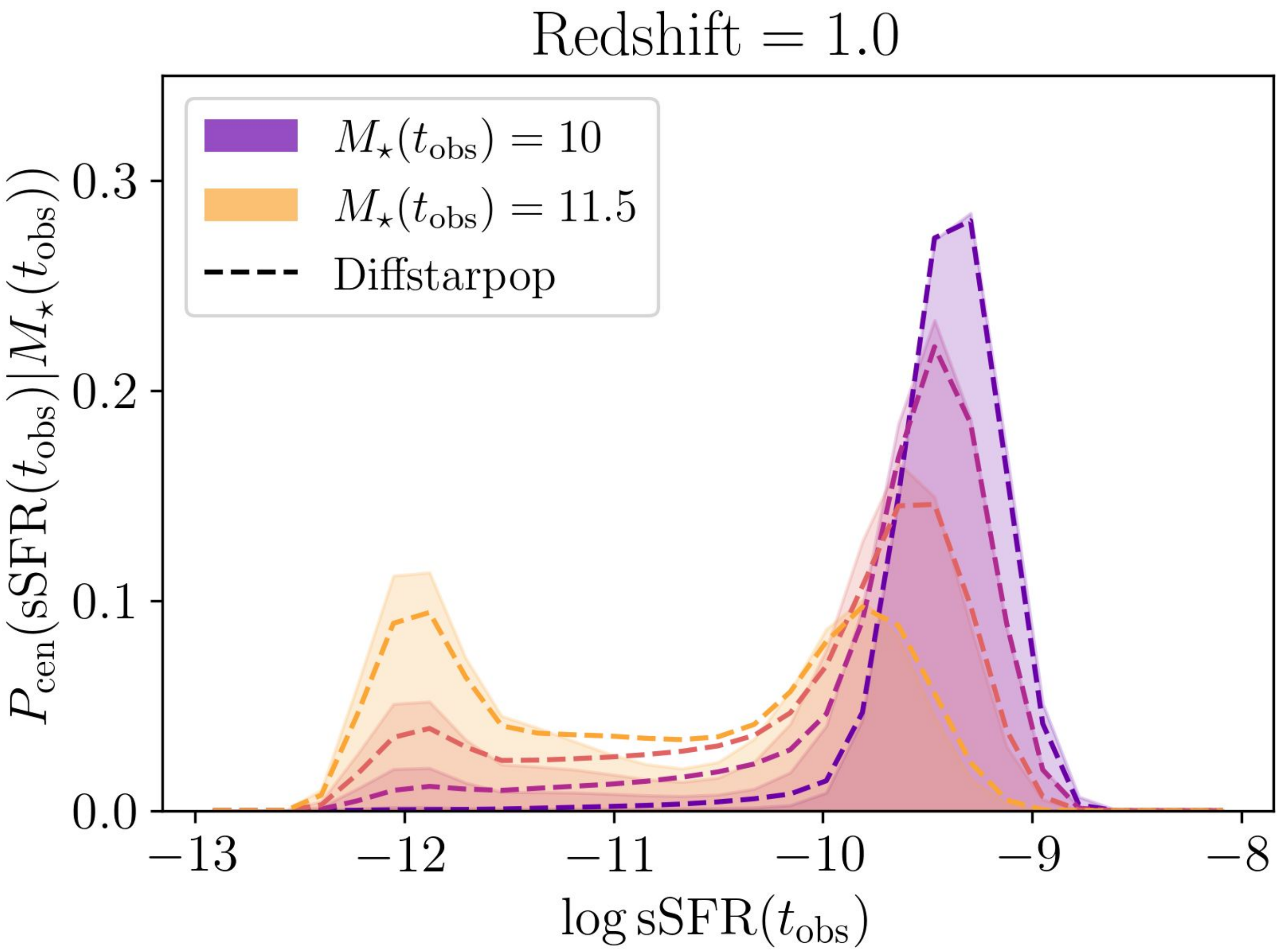
Redshift = 0.0
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
Probabilistic and differentiable prediction and selections JAX
$P_{\rm cen}(\rm sSFR(t_{obs})|M_\star(t_{obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \rm sSFR(t_{obs})$

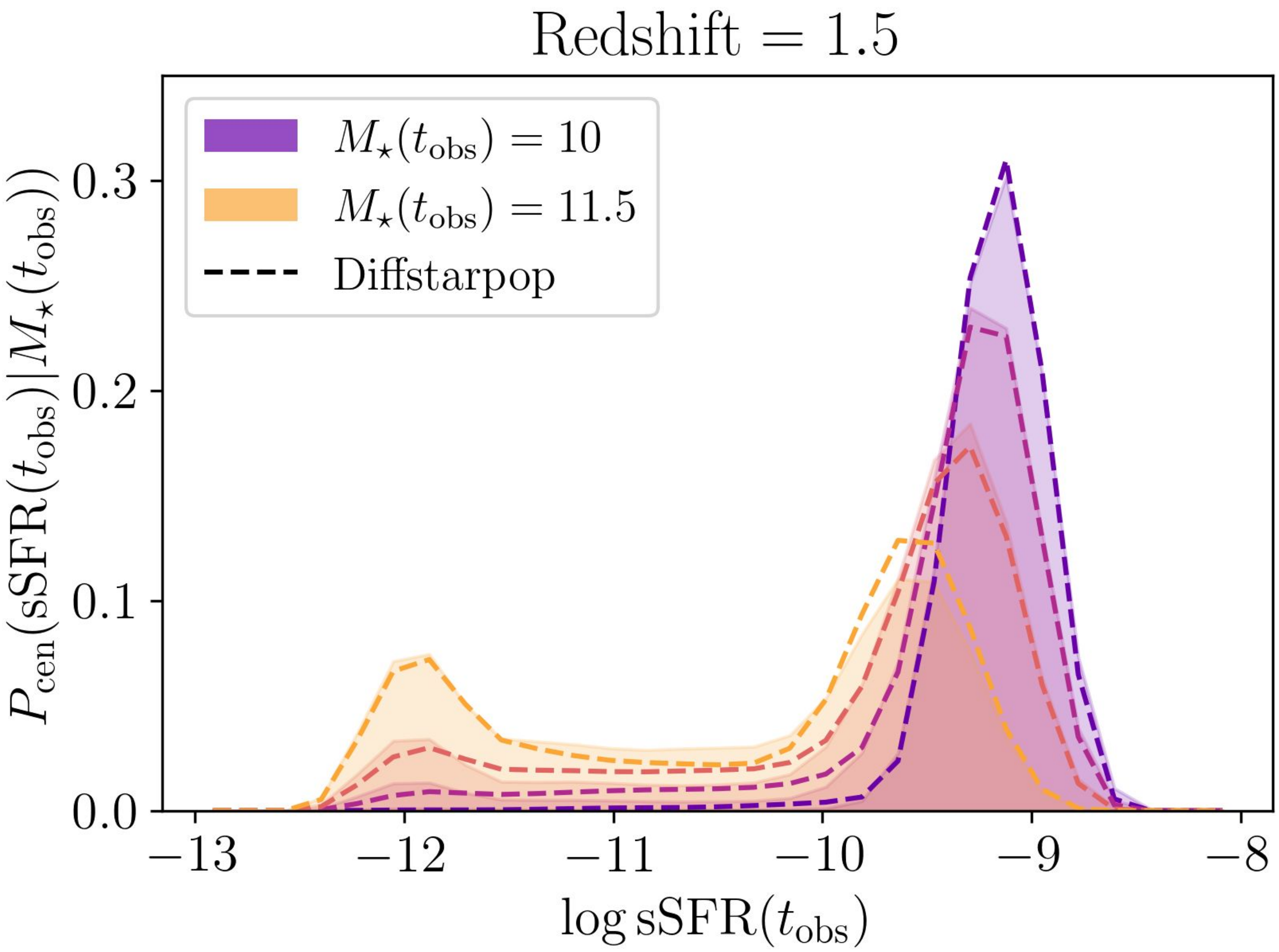


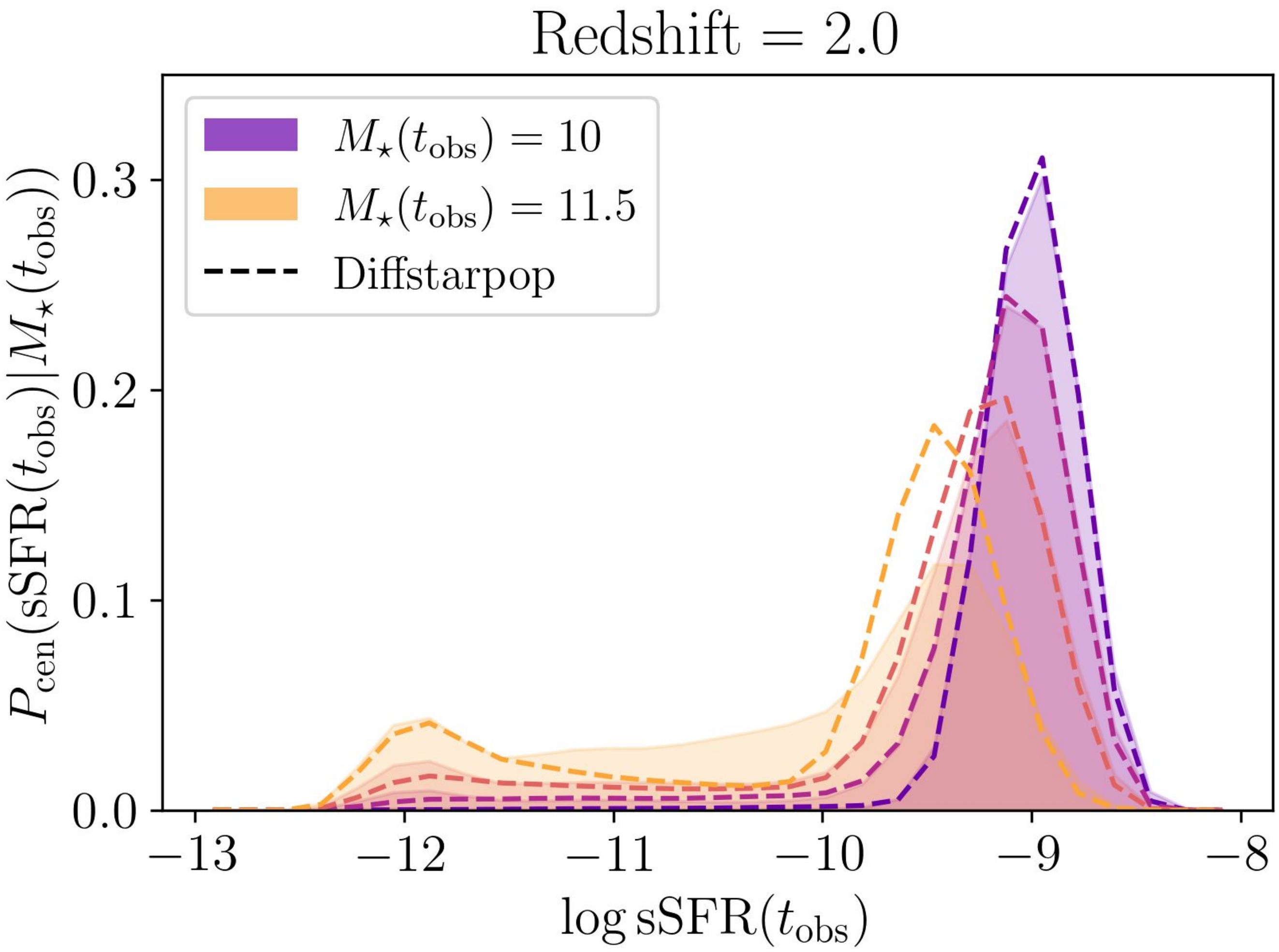
Redshift = 0.0
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
$P_{\rm cen}(\mathrm{sSFR}(t_{\rm obs})|M_\star(t_{\rm obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \mathrm{sSFR}(t_{\rm obs})$



Redshift = 0.5
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
$P_{\rm cen}(\mathrm{sSFR}(t_{\rm obs})|M_\star(t_{\rm obs}))$
$\log \mathrm{sSFR}(t_{\rm obs})$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8


Redshift = 1.0
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
$P_{\rm cen}(\rm sSFR(t_{\rm obs})|M_\star(t_{\rm obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \rm sSFR(t_{\rm obs})$


Redshift = 1.5
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
$P_{\rm cen}(\mathrm{sSFR}(t_{\rm obs})|M_\star(t_{\rm obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \mathrm{sSFR}(t_{\rm obs})$



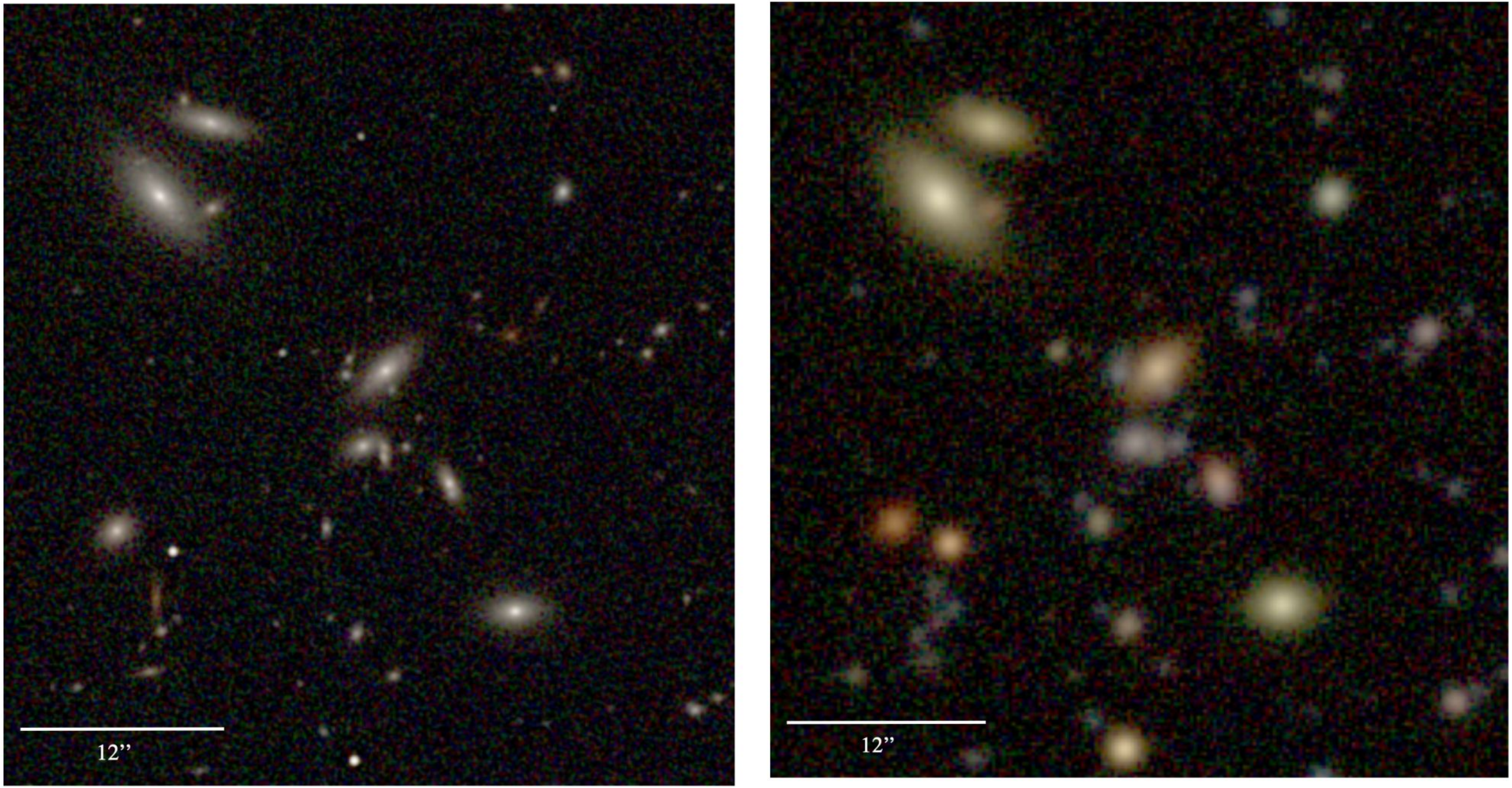
Redshift = 2.0
$M_\star(t_{\rm obs}) = 10$
$M_\star(t_{\rm obs}) = 11.5$
Diffstarpop
$P_{\rm cen}(\rm sSFR(t_{\rm obs})|M_\star(t_{\rm obs}))$
0.0
0.1
0.2
0.3
−13
−12
−11
−10
−9
−8
$\log \rm sSFR(t_{\rm obs})$

First applications using a prototype model

2501.05632
OpenUniverse
PI: Troxel

Roman + LSST image simulations (Galsim) with Outer Rim + prototype Diffsky model)

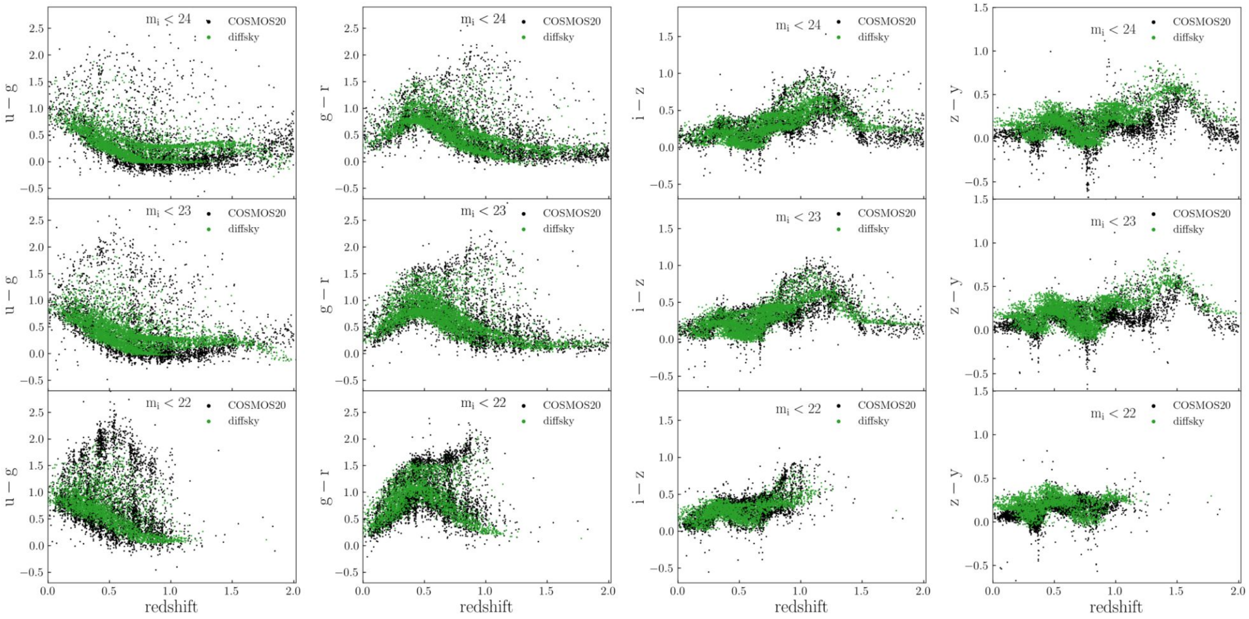
2501.05632



First applications using a prototype model

2501.05632
OpenUniverse
Data access.

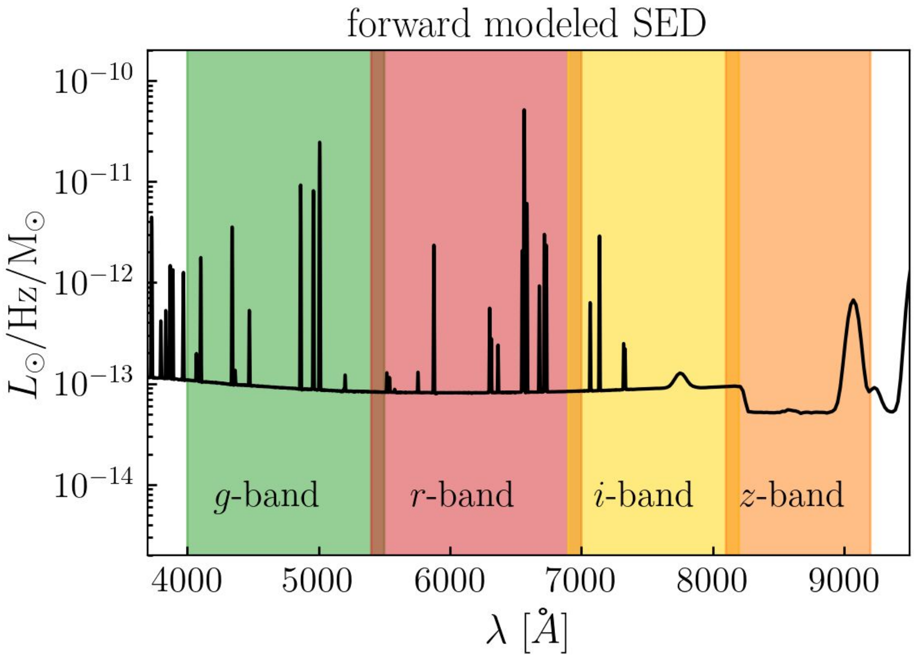
- **Prototype model:**
  - Using a prototype version of Diffsky on SMDPL.
  - Transferring galaxies from SMDPL to Outer Rim by matching on halo properties.



DSPS: Differentiable GPU-Accelerated SEDs/Photometry

DSPS:
2112.06830

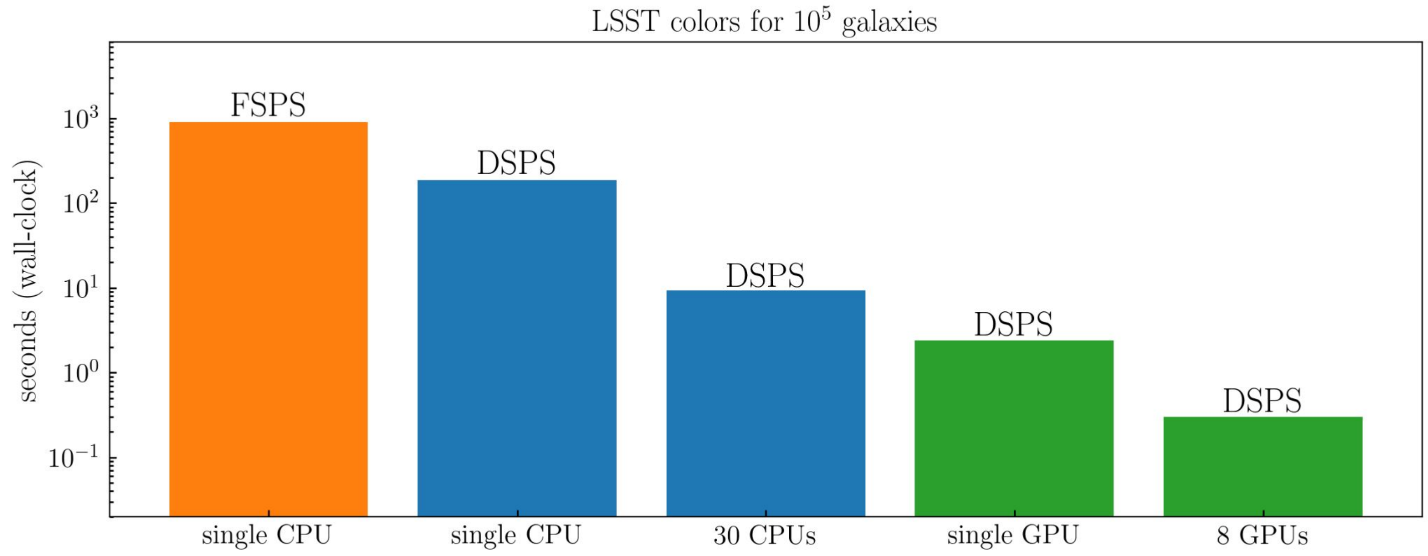
- SPS models include ingredients for dust, bursty star formation, metallicity, etc.



DSPS: Differentiable GPU-Accelerated SEDs/Photometry

DSPS:
2112.06830

- SPS models include ingredients for dust, bursty star formation, metallicity, etc.
- **Enormous performance gains** from DSPS: a JAX-based implementation of SPS



Conclusions

Long term goals:

- Multi-wavelength, same-sky observable predictions.
- Predicting Nonlinear galaxy clustering.
- Predicting Gas fraction + baryonification lensing predictions.
- Predictions for Multiple samples and their cross correlations.
- SED modeling and Redshift Calibration

Additional slides

DiffstarPop: Evolution of galaxy populations

$$\theta_{\mathrm{SFH}} \rightarrow \mathrm{SFH(t)}$$

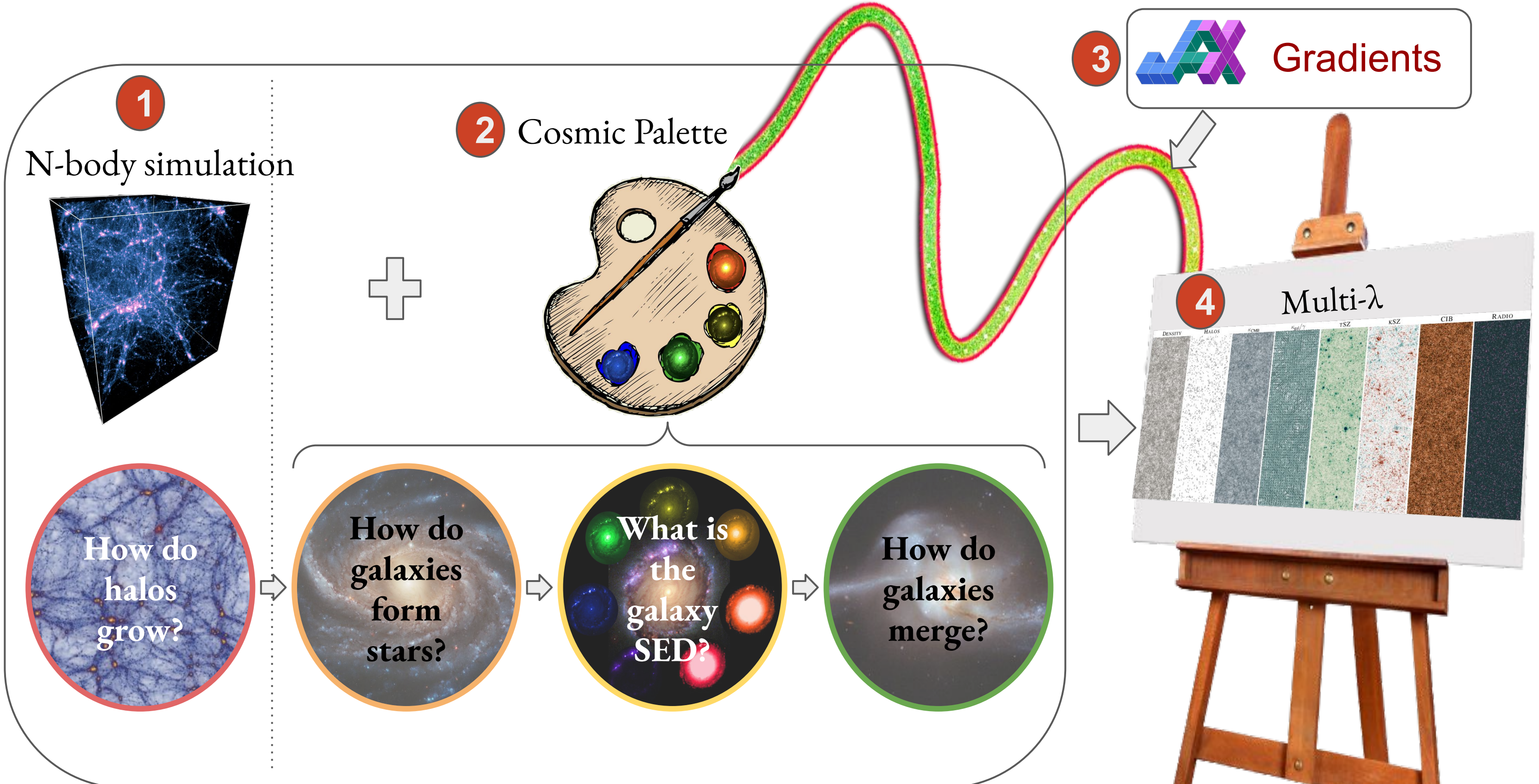
$$\Psi_{\mathrm{SFH}} \rightarrow P(\theta_{\mathrm{SFH}}|\Psi_{\mathrm{SFH}}) \rightarrow P(\mathrm{SFH(t)})$$

Mixture model:

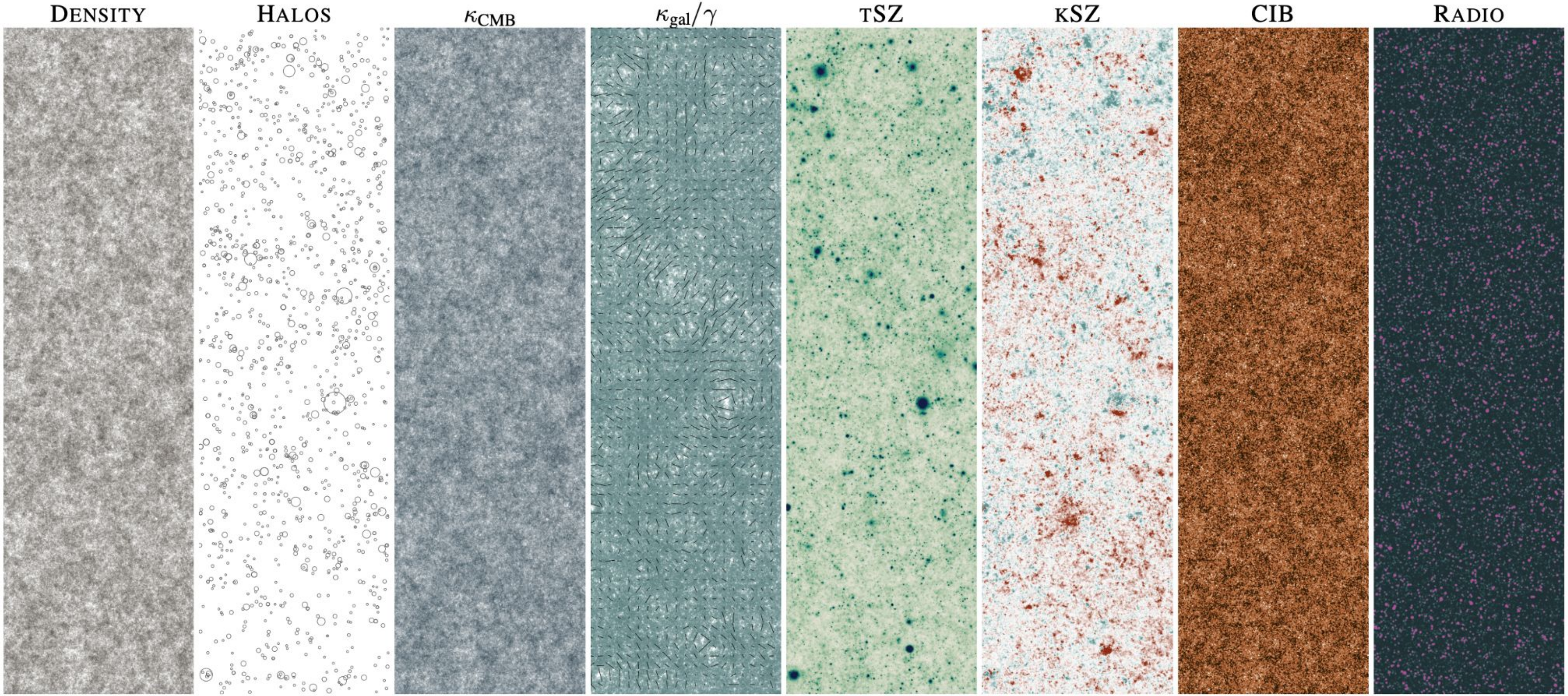
$$P(\theta_{\mathrm{SFH}}|\Psi_{\mathrm{SFH}}) = (1 - f_{\mathrm{quench}}) \overbrace{P_{\mathrm{MS}}(\theta_{\mathrm{SFH}}|\Psi_{\mathrm{MS}})}^{\text{Main sequence galaxies}} + f_{\mathrm{quench}} \overbrace{P_{\mathrm{Q}}(\theta_{\mathrm{SFH}}|\Psi_{\mathrm{Q}})}^{\text{Quenched galaxies}}$$

Linear scaling relations of each parameter with present-day halo mass M_0.

Applications: Differentiable Sky Predictions



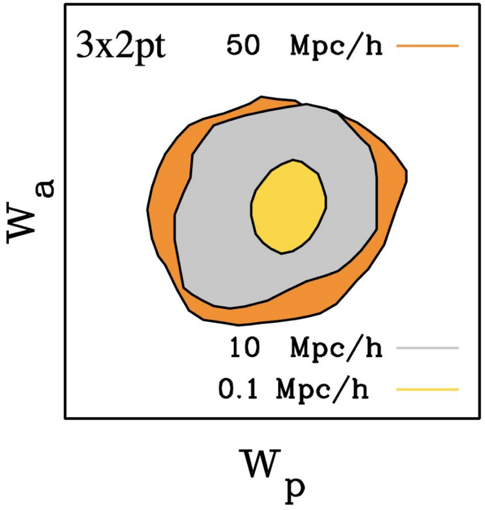
Burgeoning amount of multi-λ observations

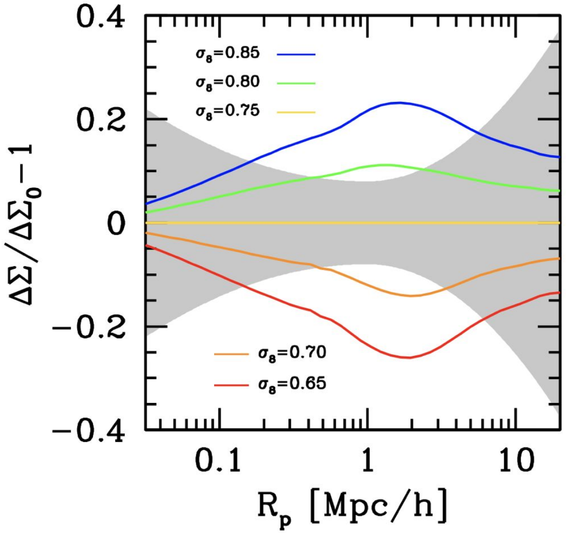


Yuuki Omori/Agora simulation

Large-scale structure in the non-linear regime

- Significant predicted cosmological gains by going to smaller scales.

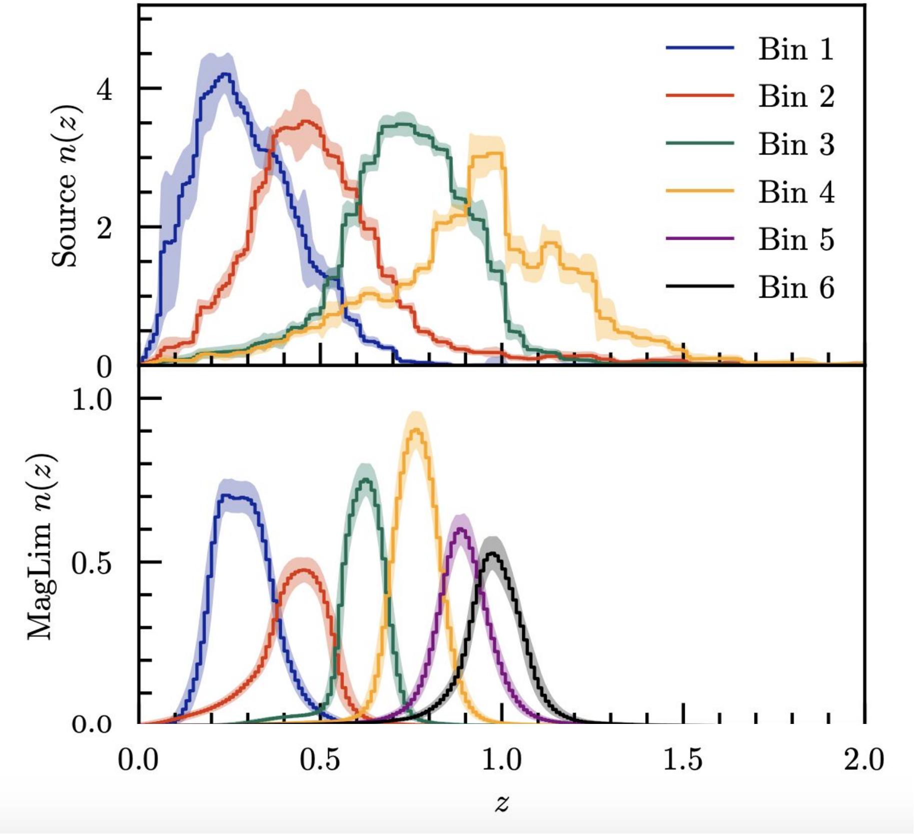




1804.03097

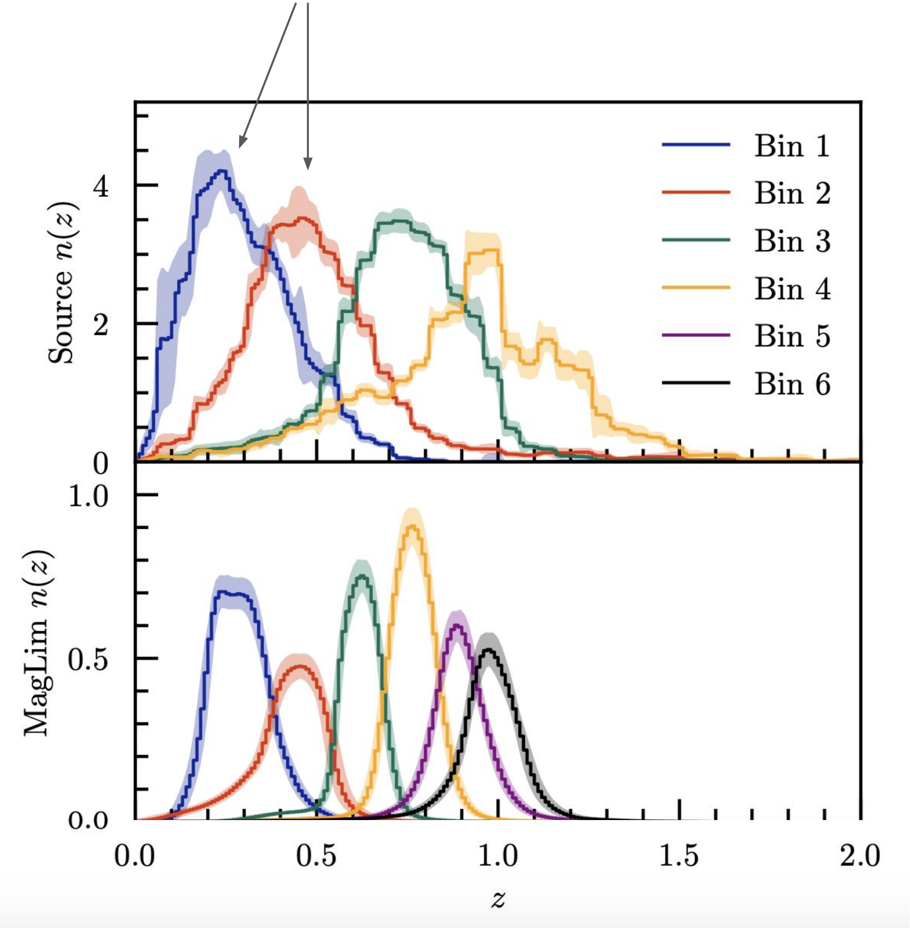
Overlap between galaxy samples

- Simultaneous forward modeling of auto- and cross-correlations of galaxy samples.
- Forward modeling of galaxy sample selections.
- Modeling the contributions of different parts of the sample to each correlation function.



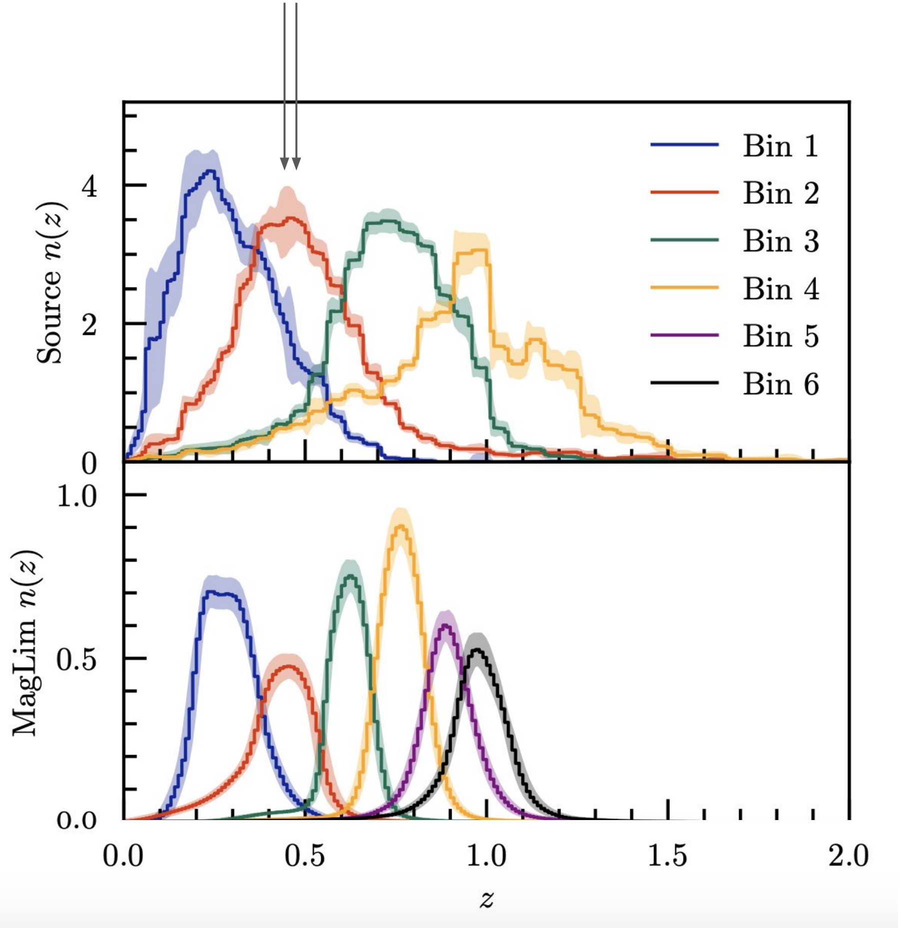
Overlap between galaxy samples

- Simultaneous forward modeling of auto- and cross-correlations of galaxy samples.
- Forward modeling of galaxy sample selections.
- Modeling the contributions of different parts of the sample to each correlation function.



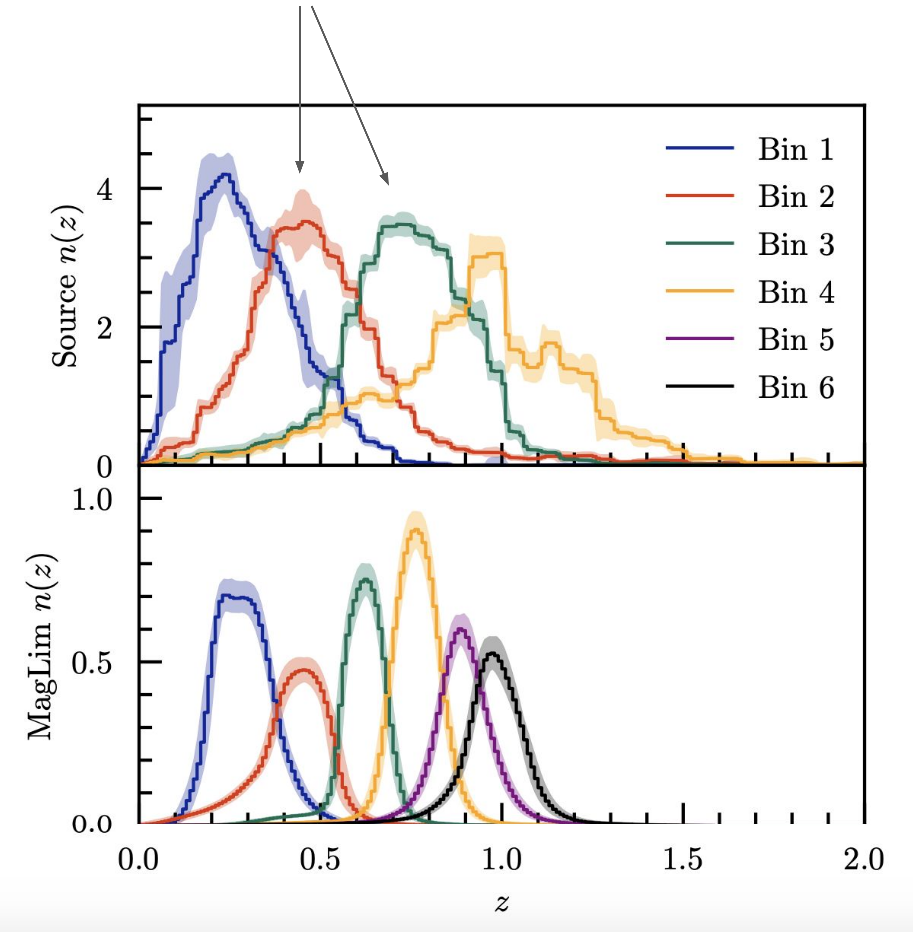
Overlap between galaxy samples

- Simultaneous forward modeling of auto- and cross-correlations of galaxy samples.
- Forward modeling of galaxy sample selections.
- Modeling the contributions of different parts of the sample to each correlation function.



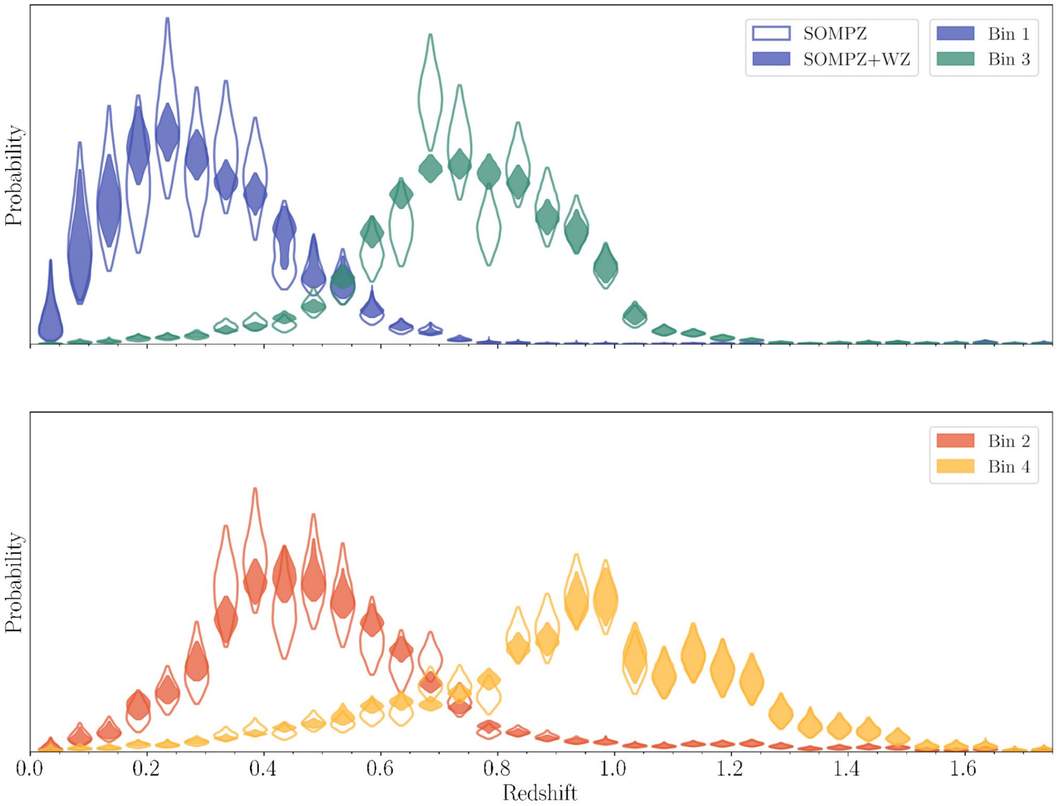
Overlap between galaxy samples

- Simultaneous forward modeling of auto- and cross-correlations of galaxy samples.
- Forward modeling of galaxy sample selections.
- Modeling the contributions of different parts of the sample to each correlation function.



Applications: Redshift calibration for Weak Lensing surveys

- Can we use the galaxy-halo connection to improve redshift calibration in photometric weak lensing surveys.
- Differentiable galaxy SED modeling.
- Physical priors inspired by models of galaxy formation.
- Hierarchical Bayesian model.
- Challenges from combining heterogeneous data (Euclid + LSST).



Applications: Redshift calibration for Weak Lensing surveys

Goal: Reducing the uncertainty in the accuracy of the mean redshift of galaxy samples.

Reality:

- DES ~ 0.013 (i~23.8)
- KiDs ~ 0.010 (i~23.5)
- HSC ~ 0.030 (i~24.5)

Upcoming requirements:

- LSST ~ 0.002 (1+z) in Y1 and 0.001 (1+z) by Y10.
- Euclid ~ 0.002 (1+z) in accuracy, 0.05(1+z) in precision.

Conclusions

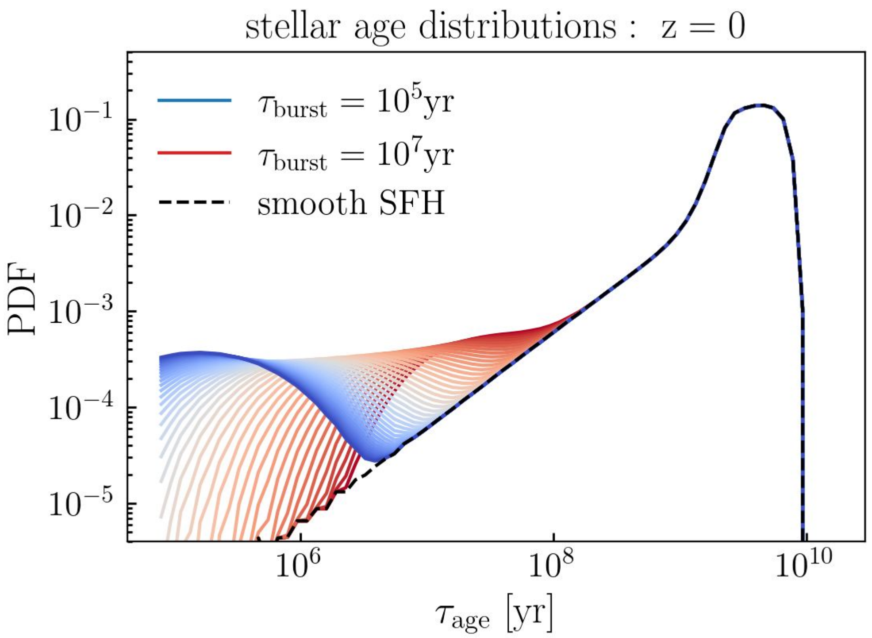
Next steps in Diffstarpop:

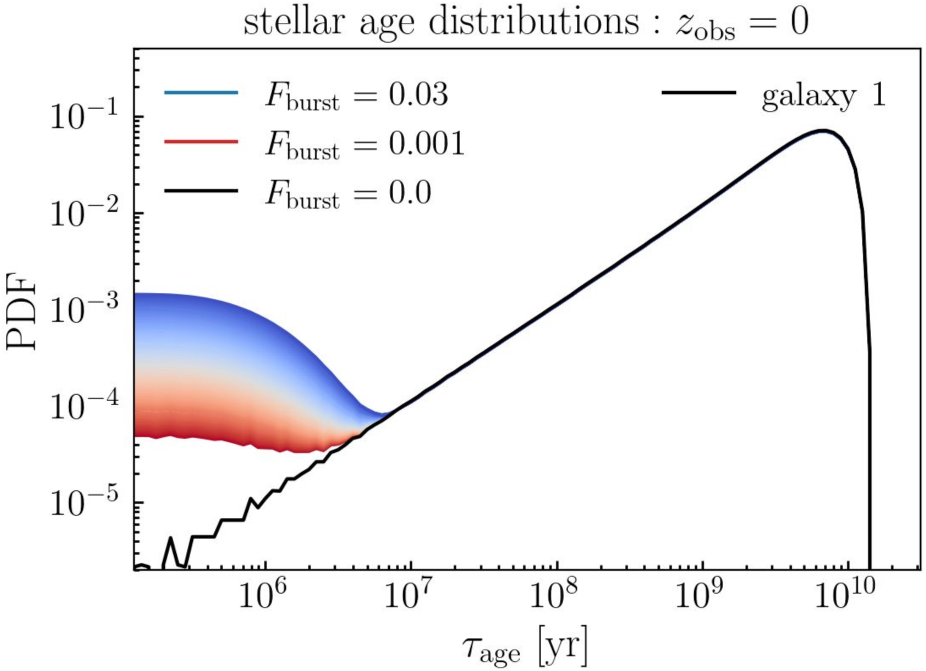
- Adding satellite specific quenching.
- Predicting galaxy assembly bias effects.
- Combining with galaxy merging: ex-situ stellar mass.

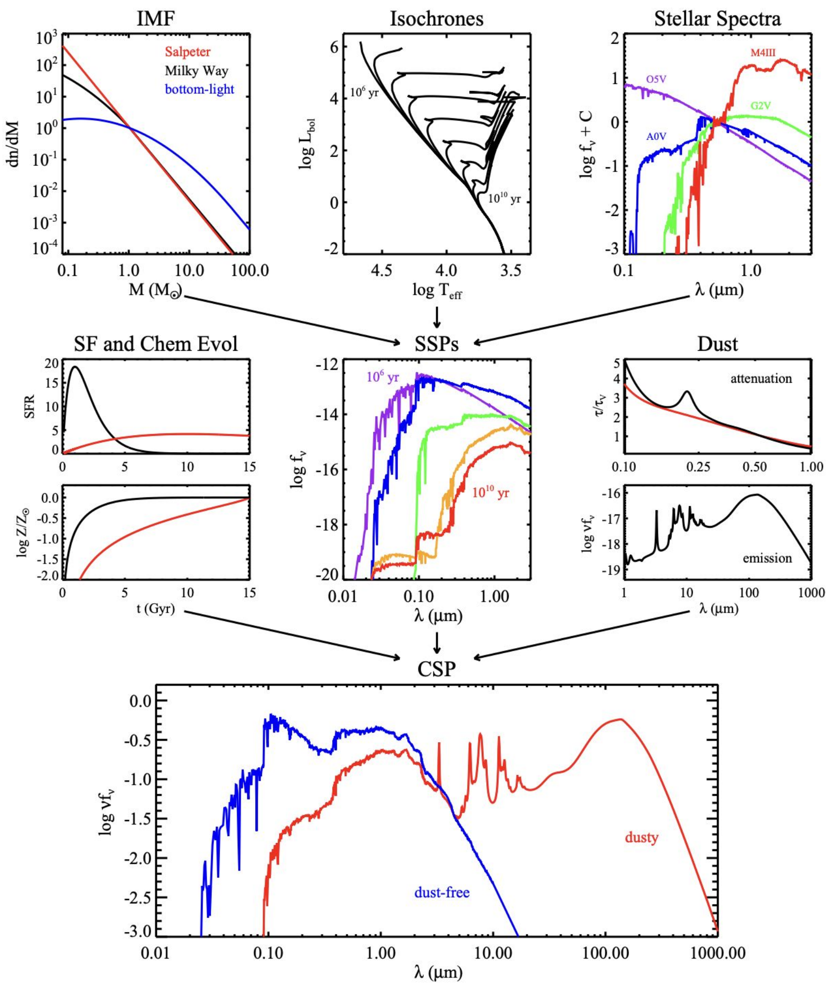
Exciting future applications:

- Multi-λ same-sky observable prediction.
- Non-linear galaxy clustering.
- Multi-sample observable prediction.
- SED modeling and Redshift Calibration

Diffburst Model of short-timescale SFH(t)







IMF
Salpeter
Milky Way
bottom-light
dn/dM
M (M⊙)
Isochrones
10^6 yr
10^10 yr
log L_bol
log T_eff
Stellar Spectra
M4III
O5V
G2V
A0V
log f_ν + C
λ (μm)
SSPs
10^6 yr
10^10 yr
log f_ν
λ (μm)
SF and Chem Evol
SFR
log Z/Z⊙
t (Gyr)
Dust
attenuation
τ/τ_V
emission
log νf_ν
λ (μm)
CSP
log νf_ν
dusty
dust-free
λ (μm)